



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 155980

**TO:** Nita M Minnifield  
**Art Unit:** 1645  
**Location:** rem/3C01/3C18  
**Serial Number:** 09/970076

**Friday, June 17, 2005**

**From:** Beverly Shears  
**Location:** Biotech-Chem Library  
**REM 1A54**  
**Phone:** 571-272-2528  
**beverly.shears@uspto.gov**

### Search Notes

#### Protein Sequence Searches – February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (uniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

*Revised*  
*MM*

**THIS PAGE BLANK (USPTO)**



XX 13-JUN-2002.  
XX 03-OCT-2001; 2001WO-US030941.  
XX 05-DEC-2000; 2000US-0251481P.  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
XX WPI; 2002-713235/77.  
XX N-PSDB; ABV73881.  
XX Novel isolated polypeptide useful for identifying agent that prevents or  
XX reduces effect of anthrax toxin on host cell, for treating human or non-  
XX human animal suffering from anthrax.  
XX Claim 1; Page 29-30; 45pp; English.  
XX  
XX The present sequence is the protein sequence of a human surface-bound  
XX anthrax toxin receptor (ATR), as predicted from an isolated cDNA clone.  
XX Anthrax toxin protective antigen (PA) binds to the ATR at a von  
XX Willebrand factor A domain located in the extracellular domain of ATR.  
XX The invention provides ATR polypeptides and polynucleotides, vectors,  
XX host cells, and transgenic and knock-out animals. It also provides  
XX methods for identifying molecules that bind the ATR and which reduce the  
XX toxicity of anthrax toxin. A claimed method for treating anthrax in a  
XX human or animal involves administering an agent that inhibits binding  
XX between PA and ATR at a level effective to reduce the severity of  
XX anthrax. Suitable agents include ATR or a PA-binding fragment of ATR, a  
XX PA-binding polypeptide at least 80% identical to these, a fusion protein,  
XX a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
XX nucleic acid  
XX  
XX Sequence 368 AA;  
XX  
XX Query Match 100.0%; Score 1536; DB 5; Length 368;  
XX Best Local Similarity 100.0%; Pred. No. 1.5e-156;  
XX Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 GGGRRRDEGGPACGCGFDLYITLDKSGSVLHHNNEIYFVEQLAHKFPISPOLRMSFYFS 60  
XX 27 GGGRRRDEGGPACGCGFDLYITLDKSGSVLHHNNEIYFVEQLAHKFPISPOLRMSFYFS 86  
XX 61 TRGTTMLKLTEDRQIRQGLBELQVLPFGSDTYMHGFERASBOIYYENRGYRTASYII 120  
XX 87 TRGTTMLKLTEDRQIRQGLBELQVLPFGSDTYMHGFERASBOIYYENRGYRTASYII 146  
XX 121 ALTDGELHEDLFFYSERERANSRDGAIYVCVGVDFNETQLARIADSKDHVPVNDGFO 180  
XX 147 ALTDGELHEDLFFYSERERANSRDGAIYVCVGVDFNETQLARIADSKDHVPVNDGFO 206  
XX 181 ALQGIHSILKSKSCIEIIAAEPSTICAGESQVVRGFGFHHANVRVLCSPKINDSVT 240  
XX 207 ALQGIHSILKSKSCIEIIAAEPSTICAGESQVVRGFGFHHANVRVLCSPKINDSVT 266  
XX 241 LNEKPFSEVDYTLCPAPILKEVGKALQVSMNDGSLFISSVITTTTHCSDS 295  
XX 267 LNEKPFSEVDYTLCPAPILKEVGKALQVSMNDGSLFISSVITTTTHCSDS 321  
XX  
XX RESULT 2  
XX AAE01439 standard; protein; 403 AA.  
XX AAE01439;  
XX 17-JUN-2001 (first entry)  
XX Human gene 4 encoded secreted protein HMLFR02, SEQ ID NO:94.  
XX Human; secreted protein; proliferative disorder; cancer; tumour;

KM foetal abnormality; developmental abnormality; haematopoietic disorder;  
KM immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
KM inflammation; allergy; neurological disorder; Alzheimer's disease;  
KM Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
KM skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
KM cardiovascular disorder; angiosenic disorder; kidney disorder;  
KM gastrointestinal disorder; pregnancy-related disorder;  
KM endocrine disorder; infection; wound healing; vulnerability; cell culture;  
KM chemotaxis; food additive; gene therapy; binding partner identification;  
KM chromosome 19.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..27  
XX Protein /label= Signal\_peptide  
XX 26..403  
XX /note= "Mature human secreted protein"  
XX  
XX WO200134626-A1.  
XX 17-MAY-2001.  
XX  
XX 01-NOV-2000; 2000WO-US030045.  
XX  
XX 05-NOV-1999; 99US-0163581P.  
XX 30-JUN-2000; 2000US-0215133P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;  
XX WPI; 2001-308778/32.  
XX N-PSDB; AAD05303.  
XX  
XX New nucleic acid molecules encoding 28 human secreted proteins for  
XX diagnosing, preventing, creating or ameliorating medical conditions and  
XX used as food additives or preservatives.  
XX  
XX Claim 11; Page 485-486; 562pp; English.  
XX  
XX AAD05300-AAD05319 represent cDNAs corresponding to 28 human secreted  
XX protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
XX AAE01514-AAE01544 represent human secreted protein fragments or variants.  
XX The genes and their secreted proteins are useful for preventing, treating  
XX or ameliorating medical conditions, e.g., by protein or gene therapy.  
XX Pathological conditions can be diagnosed by determining the amount of the  
XX new protein in a sample or by determining the presence of mutations in  
XX the new genes. Specific uses are described for each of the 28 genes,  
XX based on the tissues in which they are most highly expressed, and include  
XX developing products for the diagnosis or treatment of proliferative  
XX disorders, cancer, tumours, foetal and developmental abnormalities,  
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
XX angiosenic disorders, kidney disorders, gastrointestinal disorders,  
XX pregnancy-related disorders, endocrine disorders, and infections. The  
XX proteins can also be used to aid wound healing and epithelial cell  
XX proliferation, to prevent skin aging due to sunburn, to maintain organs  
XX before transplantation, for supporting cell culture of primary tissues,  
XX to regenerate tissues, to identify their cognate ligands or binding  
XX partners, and in chemotaxis, and can be used as a food additive or  
XX preservative to modify storage properties. Antibodies specific for a  
XX protein of the invention can be used in alleviating symptoms associated  
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
XX present sequence represents a human secreted protein of the invention  
XX  
XX Sequence 403 AA;  
XX  
XX Query Match 100.0%; Score 1536; DB 4; Length 403;



Best Local Similarity 100.0%; Pred. No. 1,8e-156;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIYFS 60  
DB 27 GGGRRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIYFS 86  
QY 61 TRGTTLMKLTEDREIQIGLEBELQKVLPGDPTVHGEFERASEQIYYENRGYRTASVII 120  
DB 87 TRGTTLMKLTEDREIQIGLEBELQKVLPGDPTVHGEFERASEQIYYENRGYRTASVII 146  
QY 121 ALTDGEIHEDLFFYSEREARNSRDGAIVYCVGVDENETOLARIADSKDHVPVNDGFQ 180  
DB 147 ALTDGEIHEDLFFYSEREARNSRDGAIVYCVGVDENETOLARIADSKDHVPVNDGFQ 206  
QY 181 ALQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRNGFRHARVNDVLCSPKINDSVT 240  
DB 207 ALQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRNGFRHARVNDVLCSPKINDSVT 266  
QY 241 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSMNDGSLFISSSVIITTHCSDDG 295  
DB 267 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSMNDGSLFISSSVIITTHCSDDG 321

RESULT 3  
ABG63874 standard; protein; 403 AA.  
XX  
AC ABG63874,  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Human albumin fusion protein #549.  
XX

KM Albumin fusion protein; therapeutic protein X; human albumin; HA;  
KM human serum albumin; HSA; cancer; reproductive disorder;  
KM digestive disorder; immune disorder; endocrine disorder;  
KM haematopoietic disorder; neural disorder; connective disorder;  
KM cytoskeletal; antiinfectivity; antiinflammatory; antitumor;  
KM immunomodulator; anti-HIV; antidiabetic; haematologic; noctropic;  
KM neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
KM osteopathic; antiaortic;  
XX

OS Homo sapiens.  
OS Synthetic.

XX WO200177137-A1.

XX 18-OCT-2001.

XX 12-APR-2001; 2001WO-US011988.

XX 12-APR-2000; 2000US-0229358P.

XX 25-APR-2000; 2000US-0199384P.

XX 21-DEC-2000; 2000US-0256931P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX ROSEN CA, HASELTINE WA;

XX WPI; 2002-010886/01.

XX New fusion protein for treating disease e.g. diabetes comprises an  
XX albumin fused to a therapeutic protein.

XX Claim 1, Page 874-875; 2102pp; English.

XX The present invention relates to albumin fusion proteins comprising a  
XX therapeutic protein X and human albumin (HA), also known as human serum  
XX albumin, HSA). The proteins are useful for treating a disease or disorder  
XX that may be mediated by therapeutic protein X. The albumin extends the  
XX shelf-life of protein X, and may increase its biological in vitro/in vivo  
XX activity. The protein is useful for treating and diagnosing disorders

CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
CC disease, ulcerative colitis), immune disorders (e.g. acquired  
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
CC schizophrenia), and connective tissue disorders (e.g. osteoporosis, arthritis).  
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX  
SQ Sequence 403 AA;

Query Match 100.0%; Score 1536; DB 5; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1,8e-156;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIYFS 60  
DB 27 GGGRRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIYFS 86  
QY 61 TRGTTLMKLTEDREIQIGLEBELQKVLPGDPTVHGEFERASEQIYYENRGYRTASVII 120  
DB 87 TRGTTLMKLTEDREIQIGLEBELQKVLPGDPTVHGEFERASEQIYYENRGYRTASVII 146  
QY 121 ALTDGEIHEDLFFYSEREARNSRDGAIVYCVGVDENETOLARIADSKDHVPVNDGFQ 180  
DB 147 ALTDGEIHEDLFFYSEREARNSRDGAIVYCVGVDENETOLARIADSKDHVPVNDGFQ 206  
QY 181 ALQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRNGFRHARVNDVLCSPKINDSVT 240  
DB 207 ALQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRNGFRHARVNDVLCSPKINDSVT 266  
QY 241 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSMNDGSLFISSSVIITTHCSDDG 295  
DB 267 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSMNDGSLFISSSVIITTHCSDDG 321

RESULT 4  
ADL77139 standard; protein; 403 AA.

XX ADL77139;

XX 20-MAY-2004 (first entry)

XX Albumin fusion protein related therapeutic protein X, SEQ ID No 621.

KM albumin fusion protein; cytostatic; antianaemic; antiaortic;  
KM antiaesthetic; anti-HIV; immunosuppressive; antiinflammatory;  
KM antiparasitic; antibacterial; osteopathic; dermatological; antitumor;  
KM immunomodulator; antiarrhythmic; cardiac; noctropic; antiparkinsonian;  
KM nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;  
KM antidiabetic; anabolic; hypertensive; vulnary; gene therapy; cancer;  
KM reproductive system disorder; therapeutic protein.

XX Unidentified.

XX US2004010134-A1.

XX 15-JAN-2004.

XX 12-APR-2001; 2001US-00833245.

XX 12-APR-2000; 2000US-0229358P.

XX 25-APR-2000; 2000US-0199384P.

XX 21-DEC-2000; 2000US-0256931P.

XX (ROSE/) ROSEN C A.

XX (HASE/) HASELTINE W A.

XX ROSEN CA, HASELTINE WA;

XX WPI; 2004-090519/09.

XX New albumin fusion proteins, useful for diagnosing, treating, preventing

PT	or ameliorating diseases or disorders e.g. cancer, anaemia, arthritis,
Pr	asthma, inflammatory bowel disease or Alzheimer's disease.
XX	
PS	Disclosure; SEQ ID NO 621; 2799p; English.
XX	
CC	The invention relates to a novel albumin fusion protein. The invention
CC	further relates to: a composition comprising the albumin fusion protein
CC	and a pharmaceutical carrier; a kit comprising the composition of the
CC	albumin fusion protein formula; a method of treating a disease or
CC	disorder in a patient comprising the step of administering the albumin
CC	fusion protein; a method of treating a patient with a disease or disorder
CC	that is modulated by Therapeutic Protein X, or its fragment or variant;
CC	a method of extending the shelf life of Therapeutic protein X, or its
CC	fragment or variant; a nucleic acid molecule comprising a polynucleotide
CC	sequence encoding the albumin fusion protein; a vector comprising the
CC	nucleic acid molecule of the albumin fusion protein; and a host cell
CC	comprising the nucleic acid molecule of the albumin fusion protein. The
CC	albumin fusion protein and its compositions have the following
CC	activities: cytostatic, antineoplastic, antithrombotic, antisthmatic, anti-
CC	HIV, immunosuppressive, antiinflammatory, antidiabetic, antibacterial,
CC	osteoprotic, dermatological, anticancer, immunomodulator, antiarrhythmic,
CC	cardiant, neurotropic, antilipemic, nephroprotective, uteroblastic,
CC	neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
CC	hypertensive, and vulnerary. The albumin fusion protein nucleic acid may
CC	be used in gene therapy to treat disorders. The albumin fusion protein is
CC	useful for diagnosing, treating, preventing or ameliorating diseases or
CC	disorders comprising indication: X. The diseases or disorders include:
CC	cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer),
CC	immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC	lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
CC	autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC	disease), reproductive system disorders (e.g. prostaticitis, inguinal
CC	hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC	leydig tumour), musculoskeletal diseases (e.g. giant cell tumours,
CC	Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC	or cachexia), cardiovascular disease (e.g. rhabdomyoma, heart disease,
CC	arrhythmia, cardiac arrest, heart valve disease, hypernatraemia or
CC	hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC	Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay
CC	-Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC	tract infections or renal disorders), neural or sensory disease (e.g.
CC	Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC	cerebellar ataxia, attention deficit disorder, autism or obsessive
CC	compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC	occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC	disease or glomerulonephritis), digestive diseases (e.g. portal
CC	hypertension, irritable bowel disease, gastric atrophy or paracatitls)
CC	or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC	scleroderma, wound healing or epidermolysis bullosa). This sequence
CC	represents a therapeutic protein X relating to the albumin fusion protein
CC	of the invention. The sequence listing data for this specification was
CC	downloaded from the USPTO website.
CC	
XX	
XX	
Seq	Sequence 403 AA;
Query Match	100.0%; Score 1536; DB 8; Length 403;
Beef Local Similarity	100.0%; Pred. No. 1.8e-156;
Matches 295; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
1	GCGGRREDGPGACGAGFDLYFLIDSGSVLHHMNEIYYFVEQLAKHFLSPQLRMSPIYFS 60
27	GCGGRREDGPGACGAGFDLYFLIDSGSVLHHMNEIYYFVEQLAKHFLSPQLRMSPIYFS 86
61	TGRTTLMKLTERRRQIRGLBELQVLPGGDYYMHGFFRASEQIYYENRQGYRTASVYI 120
87	TGRTTLMKLTERRRQIRGLBELQVLPGGDYYMHGFFRASEQIYYENRQGYRTASVYI 146
121	ALTDELHEHDLFFYFSREANRRDICALIYYCVGVDPENETOLARLADSNDHFVPNDGFO 180
147	ALTDELHEHDLFFYFSREANRRDICALIYYCVGVDPENETOLARLADSNDHFVPNDGFO 206
161	ALQGIHSILKSKCIEIILAAEPTTCAGSFQVYVVRGNGFRHARVNDVRLCSFKINDSVT 240

Dd		207	ALQGIHSILKKSCTIELLAEPSTTCAGSEFQQVYVRNGRHNANDRVLCSPFINDSVT	266
Oy		241	LNEKFPSVEDTYLLCPAPILKEVGAKAOLGVSMNDGLSFISSSVIITTHCSDGS	295
Dd		267	LNEXFPSEDPTYLCCPAPILKEVGAKAOLGVSMNDGSLFISSSVIITTHCSDDGS	321
RESULT 5				
ID	ABB90750			
XX	ABB90750 standard; protein; 564 AA.			
AC	ABB90750;			
DT	30-MAY-2002 (first entry)			
DE				
XX	Human Tumour Endothelial Marker polypeptide SEQ ID NO 232.			
KW				
KM	Human; mouse; rat; TEM: tumour endothelial marker; NEM; PEM: cytotattic;			
KX	normal endothelial marker; pan-endothelial marker; immunostimulant;			
KW	antiangiogenic; tumour neoangiogenesis; vascularised tumor;			
KM	polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;			
KX	psoirais.			
XX				
OS	Homo sapiens.			
PN	WO200210217-A2.			
PD	07-FEB-2002.			
PF	01-AUG-2001; 2001MO-USO24031.			
PR	02-AUG-2000; 2000US-0222599P.			
PR	11-AUG-2000; 2000US-0224360P.			
PR	11-APR-2001; 2001US-0282850P.			
PA	(UYJO ) UNIV JOHNS HOPKINS.			
PI	St Croix B, Kinzler KM, Vogelstein B;			
DR	N-PDDB; ABL92104.			
XX				
PT	An isolated molecule comprising an antibody variable region which			
PT	specially binds to an extracellular domain of a tumor endothelial			
PT	marker (TEM) protein, useful for inhibiting tumor growth.			
PS	Claim 1; Page 209-210; 331pp; English.			
XX				
CC	The invention relates to an isolated molecule comprising an antibody			
CC	variable region which specifically binds to an extracellular domain of a			
CC	tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,			
CC	ABBR0749, ABB90750 and ABB90769. The antibodies which bind to TEM			
CC	proteins have cytotoxic, immunostimulant and antiangiogenic activity.			
CC	They are useful for inhibiting tumour growth, neoangiogenesis in subjects			
CC	bearing a vascularised tumour, polycystic kidney disease, diabetic			
CC	retinopathy, Rheumatoid arthritis and psoriasis. Human, mouse and rat TEM			
CC	genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)			
CC	are disclosed, as are marker oligonucleotide sequences: tumour			
CC	endothelial markers (NEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal			
CC	endothelial markers (PEM) ABL92042-ABL92074; and pan-endothelial markers			
CC	(PEM) ABL91903-ABL91995			
SQ	Sequence 564 AA;			
Query Match	100.0%; Score 1536; DB 5; Length 564;			
Best Local Similarity	100.0%; Pred. No. 3e-156;			
Matches 295; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
Oy		1	GCGGRREDCGGACGCFDLFYFLDKSGSVLAHNMIEIFYVFQLAHKFTPOLRMSTFFVS	60
Dd		27	GCGGRREDGCACGTGGFDLYTLIDKSSGLVAHMMEITYFVDLHKFTIPOLRMSTFFVS	86
Oy		61	TGTTTLMKTEDRBQIRGLELEOKVLPFGDJTNMHGEFRASEOIYYENRCGYRTASVII	120

Db 87 TRGTTLMKLTEDRQIRQGLBELOKVLPGSDTYMHGEFERSSEQIYYENRQGYRTASVYI 146  
 Qy 121 ALTDGELHEDLFFYSEREARNRDLGAIYVCVGYDNEFETOLARIADSKDHVPVNDGFO 180  
 Db 147 ALTDGELHEDLFFYSEREARNRDLGAIYVCVGYDNEFETOLARIADSKDHVPVNDGFO 206  
 Qy 181 ALGGIHSILKSCIEIIAABPSTTCAGESFOVVRNGFRHARNVDRVLCSPKINDSVT 240  
 Db 207 ALGGIHSILKSCIEIIAABPSTTCAGESFOVVRNGFRHARNVDRVLCSPKINDSVT 266  
 Qy 241 LNEKPSVEDTYLLCPAPILKEVGKALQVSNMDGLSFSSSVITTTTCHSDGS 295  
 Db 267 LNEKPSVEDTYLLCPAPILKEVGKALQVSNMDGLSFSSSVITTTTCHSDGS 321

RESULT 6  
 ABB90724  
 ID ABB90724 standard; protein; 564 AA.  
 AC ABB90724;  
 XX  
 DT 30-MAY-2002 (first entry)  
 XX  
 DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 187.  
 XX  
 KW Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostratic;  
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;  
 KW antiangiogenic; tumour; neovascularisation; vascularised tumour;  
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;  
 KW psoriasis.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200210217-A2.  
 PD 07-FEB-2002.  
 XX  
 PF 01-AUG-2001; 2001WO-US024031.  
 PR 02-AUG-2000; 2000US-0222599P.  
 PR 11-AUG-2000; 2000US-0224360P.  
 PR 11-APR-2001; 2001US-0282850P.  
 XX  
 PA (UWJO ) UNIV JOHNS HOPKINS.  
 XX  
 PI St Croix B, Kinzler KW, Vogelstein B;  
 XX  
 DR WPI; 2002-291856/33.  
 XX  
 PT An isolated molecule comprising an antibody variable region which  
 PT specifically binds to an extracellular domain of a tumor endothelial  
 PT marker (TEM) protein, useful for inhibiting tumor growth.  
 XX  
 PS Disclosure; Page 136-137; 33pp; English.  
 XX  
 CC The invention relates to an isolated molecule comprising an antibody  
 CC variable region which specifically binds to an extracellular domain of a  
 CC tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740,  
 CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM  
 CC proteins have cytostratic, immunostimulant and antiangiogenic activity.  
 CC They are useful for inhibiting tumor growth, neovascularisation in subjects  
 CC bearing a vascularised tumour, polycystic kidney disease, diabetic  
 CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM  
 CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)  
 CC are disclosed, as are marker oligonucleotide sequences: tumour  
 CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal  
 CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers  
 CC (PEM) ABL91903-ABL91995  
 XX  
 PS Sequence 564 AA;  
 SQ

Query Match 100.0%; Score 1536; DB 5; Length 564;

Best Local Similarity 100.0%; Pred. No. 3e-156;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GGGRRDDGPAACGAGFDLFIIDKSGSVLHHNNEIYYFVQLAHKFISPOLRMSFIYFS 60  
 Db 27 GGGRRDDGPAACGAGFDLFIIDKSGSVLHHNNEIYYFVQLAHKFISPOLRMSFIYFS 86  
 Qy 61 TRGTTLMKLTEDRQIRQGLBELOKVLPGSDTYMHGEFERSSEQIYYENRQGYRTASVYI 120  
 Db 87 TRGTTLMKLTEDRQIRQGLBELOKVLPGSDTYMHGEFERSSEQIYYENRQGYRTASVYI 146  
 Qy 121 ALTDGELHEDLFFYSEREARNRDLGAIYVCVGYDNEFETOLARIADSKDHVPVNDGFO 180  
 Db 147 ALTDGELHEDLFFYSEREARNRDLGAIYVCVGYDNEFETOLARIADSKDHVPVNDGFO 206  
 Qy 181 ALGGIHSILKSCIEIIAABPSTTCAGESFOVVRNGFRHARNVDRVLCSPKINDSVT 240  
 Db 207 ALGGIHSILKSCIEIIAABPSTTCAGESFOVVRNGFRHARNVDRVLCSPKINDSVT 266  
 Qy 241 LNEKPSVEDTYLLCPAPILKEVGKALQVSNMDGLSFSSSVITTTTCHSDGS 295  
 Db 267 LNEKPSVEDTYLLCPAPILKEVGKALQVSNMDGLSFSSSVITTTTCHSDGS 321

RESULT 7  
 ABB54904  
 ID ABB54904 standard; protein; 564 AA.  
 AC ABB54904;  
 XX  
 DT 08-JAN-2003 (first entry)  
 XX  
 DE Human anthrax toxin receptor.  
 XX  
 KW Anthrax; toxin; receptor; human; TEM8; antibacterial.  
 KW  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200246228-A2.  
 PD 13-JUN-2002.  
 XX  
 PF 03-OCT-2001; 2001WO-US030941.  
 PR 05-DEC-2000; 2000US-0251481P.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
 XX  
 DR WPI; 2002-713235/77.  
 DR N-PSDB; ABV73882.  
 XX  
 PT Novel isolated polypeptide useful for identifying agent that prevents or  
 PT reduces effect of anthrax toxin on host cell, for treating human or non-  
 PT human animal suffering from anthrax.  
 XX  
 PS Claim 1; Page 37-39; 45pp; English.  
 XX

CC The present sequence is the protein sequence of a human surface-bound  
 CC anthrax toxin receptor (ATR), previously designated TEM8, and identified  
 CC following a database screening using a newly isolated human surface-bound  
 CC ATR (see ABP54903). The present sequence differs from the newly isolated  
 CC human ATR only in the cytoplasmic domain (221 rather than 25 amino acids  
 CC long), suggesting differential splicing of a primary mRNA transcript.  
 CC Anthrax toxin protective antigen (PA) binds to these ATRs at a von  
 CC Willebrand factor A domain located in the extracellular domain of ATR.  
 CC The invention provides ATR polypeptides and polynucleotides, vectors,  
 CC host cells, and transgenic and knock-out animals. It also provides the  
 CC methods for identifying molecules that bind the ATR and which reduce the  
 CC toxicity of anthrax toxin. A claimed method for treating anthrax in a  
 CC human or animal involves administering an agent that inhibits binding  
 CC between PA and ATR at a level effective to reduce the severity of  
 CC anthrax. Suitable agents include the present polypeptide or a PA-binding  
 CC fragment of it, a PA-binding polypeptide at least 80% identical to these,  
 CC a fusion protein, a monoclonal or polyclonal antibody, a polysaccharide,  
 CC a lipid or a nucleic acid

XX Sequence 564 AA:

Query Match 100.0%; Score 1536; DB 5; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 3e-156;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGGPACGCGFDLYFILDKSGSVLHNNHNIYFVYEQLAHKFISPOLKMSFIVFS 60  
 DB 27 GGGRRREDGGPACGCGFDLYFILDKSGSVLHNNHNIYFVYEQLAHKFISPOLKMSFIVFS 86  
 QY 61 TRGTTLMKLTEDREIQROGLEBLQKVLPGDVTMHGFERASQIYYENRQGYRTASVIT 120  
 DB 87 TRGTTLMKLTEDREIQROGLEBLQKVLPGDVTMHGFERASQIYYENRQGYRTASVIT 146  
 QY 121 ALTDGELHEDLPFYSEERANRSRDIGAIVYCVGVKDFNETQLARIADSKDHVFPVNDGFO 180  
 DB 147 ALTDGELHEDLPFYSEERANRSRDIGAIVYCVGVKDFNETQLARIADSKDHVFPVNDGFO 206  
 QY 181 ALOGIHSILKSCIEILAEPSITICAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 240  
 DB 207 ALOGIHSILKSCIEILAEPSITICAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 266  
 QY 241 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDGS 295  
 DB 267 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDGS 321

RESULT 8  
 ABUS4457  
 ID ABUS4457 standard; protein; 564 AA.

XX AC ABUS4457;  
 XX DT 12-MAR-2003 (first entry)

XX DE Human tumour endothelial marker TEM 19.

XX KW Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;  
 KW Tumour endothelial marker; normal endothelial marker; PEM;  
 KW pan-endothelial marker; polycystic kidney disease; psoriasis;  
 KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;  
 KW neovascularization; immune response; cytostatic; antidiabetic;  
 KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic.

XX OS Homo sapiens.

XX PN WO200283874-A2.

XX PD 24-OCT-2002.

XX PF 10-APR-2002; 2002WO-US008253.

XX PR 11-APR-2001; 2001US-0282850P.

XX PR 06-FEB-2002; 2002US-0354262P.

XX (UYJO) UNIV JOHNS HOPKINS.  
 PA Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;  
 XX WPI, 2003-093016/08.  
 XX N-PSDB; ABX72029.  
 XX DR N-PSDB; ABX72029.  
 XX PT New purified human transmembrane protein, designated as tumor endothelial  
 PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors,  
 PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or  
 PT psoriasis.

XX Disclosure; Page 226-227; 374pp; English.

XX The present invention relates to a novel method for the isolation of  
 CC endothelial cells (ECs), and the identification of genes expressed in  
 CC normal and tumour ECs. Tumour endothelial marker (TEM) normal  
 CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are  
 CC identified in human ECs. The human EC marker proteins and the  
 CC polynucleotide sequences encoding them are useful for detecting,  
 CC diagnosing or treating tumours as well as polycystic kidney disease,  
 CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also  
 CC useful for inhibiting neovascularization or tumour angiogenesis, for  
 CC inducing an immune response to tumour endothelial cells in a patient, or  
 CC for identifying candidate drugs for treating tumours. The present  
 CC sequence represents a human TEM or NEM protein of the invention

XX Sequence 564 AA:

Query Match 100.0%; Score 1536; DB 6; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 3e-156;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGGPACGCGFDLYFILDKSGSVLHNNHNIYFVYEQLAHKFISPOLKMSFIVFS 60  
 DB 27 GGGRRREDGGPACGCGFDLYFILDKSGSVLHNNHNIYFVYEQLAHKFISPOLKMSFIVFS 86  
 QY 61 TRGTTLMKLTEDREIQROGLEBLQKVLPGDVTMHGFERASQIYYENRQGYRTASVIT 120  
 DB 87 TRGTTLMKLTEDREIQROGLEBLQKVLPGDVTMHGFERASQIYYENRQGYRTASVIT 146  
 QY 121 ALTDGELHEDLPFYSEERANRSRDIGAIVYCVGVKDFNETQLARIADSKDHVFPVNDGFO 180  
 DB 147 ALTDGELHEDLPFYSEERANRSRDIGAIVYCVGVKDFNETQLARIADSKDHVFPVNDGFO 206  
 QY 181 ALOGIHSILKSCIEILAEPSITICAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 240  
 DB 207 ALOGIHSILKSCIEILAEPSITICAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 266  
 QY 241 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDGS 295  
 DB 267 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDGS 321

RESULT 9  
 ABUS4431  
 ID ABUS4431 standard; protein; 564 AA.

XX AC ABUS4431;  
 XX DT 12-MAR-2003 (first entry)

XX DE Human tumour endothelial marker TEM 8.

XX KW Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;  
 KW Tumour endothelial marker; normal endothelial marker; PEM;  
 KW pan-endothelial marker; polycystic kidney disease; psoriasis;  
 KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;  
 KW neovascularization; immune response; cytostatic; antidiabetic;  
 KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic.

XX OS Homo sapiens.

XX	WM0200283874-A2.
PN	
PD	24-OCT-2002.
XX	
PF	10-APR-2002; 2002MO-US008253.
PR	
XX	
PR	11-APR-2001; 2001US-0282850P.
XX	
PR	06-FEB-2002; 2002US-0354262P.
PA	(UWJO ) UNIV JOHNS HOPKINS.
PI	Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;
DR	WPI, 2003-093016/08.
PT	New purified human transmembrane protein, designated as tumor endothelial
PT	marker (TEM) 3, useful for detecting, diagnosing or treating tumors, or
PT	polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or
PS	psoriasis.
XX	
PS	Diclosure; Page 138-139; 374pp; English.
XX	
CC	The present invention relates to a novel method for the isolation of
CC	endothelial cells (ECs), and the identification of genes expressed in
CC	normal and tumour ECs. Tumour endothelial marker (TEM), normal
CC	endothelial marker (NEM), and pan-endothelial marker (PEM) genes are
CC	identified in human ECs. The human EC marker proteins and the
CC	polynucleotide sequences encoding them are useful for detecting,
CC	diagnosing or treating tumours as well as polycystic kidney disease,
CC	diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also
CC	useful for inhibiting neovascularogenesis or tumour angiogenesis, for
CC	inducing an immune response to tumour endothelial cells in a patient, or
CC	for identifying candidate drugs for treating tumours. The present
CC	sequence represents a human TEM or NEM protein of the invention
XX	
SQ	Sequence 564 AA;
	Query Match 100.0%; Score 1536; DB 6; Length 564;
	Best Local Similarity 100.0%; Pred. No. 3e-156;
	Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 GGGRRDDGGPACVCGFDLYFLIDSGSVLHHMNEIYYFVEQLAKHFISPOLRMSFIYS 60
DB	27 GGGRREDDGGPACVCGFDLYFLIDSGSVLHHMNEIYYFVEQLAKHFISPOLRMSFIYS 86
QY	61 TRGTTLMKLTDREDRIROGLEELQVYLPGGDITTMHBGFERRASEQIYYENRGYRTASVII 120
DB	87 TRGTTLMKLTDREDRIROGLEELQVYLPGGDITTMHBGFERRASEQIYYENRGYRTASVII 146
QY	121 ALTQGLHEHDLPFYSEERANRSRDIGAIYYCVGVDDFNMTOLARIADSDNHVPNNDFQ 180
DB	147 ALTQGLHEHDLPFYSEERANRSRDIGAIYYCVGVDDFNMTOLARIADSDNHVPNNDFQ 206
QY	181 ALQGIIHSILKKSCIEILAAEPSTICAGSPVVVRNGNGFRARAVNDRLCSFKINDSYT 240
DB	207 ALQGIIHSILKKSCIEILAAEPSTICAGSPVVVRNGNGFRARAVNDRLCSFKINDSYT 266
QY	241 LNEKRFVSVDITYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTTTCSDGS 295
DB	267 LNEKRFVSVDITYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTTTCSDGS 321
RESULT 10	
ADJ70017	
ID	ADJ70017 standard; protein; 564 AA.
XX	
AC	ADJ70017;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Human heat mitochondrial protein as a therapeutic target SeqID1823.

KM	mitochondrial, human; screening assay; diabetes mellitus;
KM	Huntington's disease; osteoarthritis;
KM	Leber's hereditary optic neuropathy; LHON;
KM	mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KM	myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KM	neuroprotective; neurotropic; antidiabetic; anticonvulsant; antiarthritic;
KM	osteopathic; ophthalmological; cytoskeletal.
OS	Homo sapiens.
XX	WO2003087768-A2.
XX	23-OCT-2003.
XX	04-APR-2003; 2003WO-US010870.
XX	12-APR-2002; 2002US-0372643P.
PR	17-JUN-2002; 2002US-0398987P.
PR	20-SEP-2002; 2002US-0412418P.
XX	(MITO-) MITOKOR.
PA	(BUCK-) BUCK INST AGE RES.
XX	Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI	Warrock DE;
PI	WPI, 2003-845369/78.
XX	Identifying a mitochondrial target for drug screening assays and for
XX	treating diseases associated with altered mitochondrial function,
PT	PT comprises detecting a modified polypeptide in a sample and correlating
PT	with the disease.
PS	Claim 1; SEQ ID NO 1823; 180pp; English.
XX	This invention relates to novel mitochondrial targets that can be used
CC	for therapeutic intervention in treating a disease associated with
CC	altered mitochondrial function. Specifically, it refers to a method for
CC	identifying proteins of the human heart mitochondrial proteome that are
CC	useful for drug screening assays, as well as therapeutic targets. The
CC	present invention describes a method for identifying such proteins that
CC	can be used in the treatment of various diseases associated with altered
CC	mitochondrial function including diabetes mellitus, Huntington's disease,
CC	osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC	encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC	ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC	compositions have neuroprotective, neurotropic, antidiabetic,
CC	anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC	cytoskeletal activities. This polypeptide sequence is a human heart
CC	mitochondrial protein of the invention.
XX	Sequence 564 AA;
SO	Query Match 100.0%; Score 1536; DB 7; Length 564;
	Best Local Similarity 100.0%; Pred. No. 3e-156;
	Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 GCGGRRDDGGPACGCGFDLYFLIDKSGSYLHMNTEYFVEQLAKFTISPQLRMSFTIVS 60
DB	27 GCGGRRDDGGPACGCGFDLYFLIDKSGSYLHMNTEYFVEQLAKFTISPQLRMSFTIVS 86
QY	61 TRGTTMLMCLTDRREIRIGGLELQVLRGGDVTYMHGPFRRASEQIYYENRGGRTASVII 120
DB	87 TRGTTMLMCLTERREIRIGGLELQVLRGGDVTYMHGPFRRASEQIYYENRGGRTASVII 146
QY	121 ALVTGELHEDLFFYSERREANRSRDGAIIVYCVGVDDFNETQLARITADSKDHFVPVNDGFO 180
DB	147 ALVTGELHEDLFFYSERREANRSRDGAIIVYCVGVDDFNETQLARITADSKDHFVPVNDGFO 206
QY	181 ALGGIHSILKKSCEIILAAEPTTCAGESFOVVVRNGGFRHARNDVRLVCSFKINDSVT 240
DB	207 ALGGIHSILKKSCEIILAAEPTTCAGESFOVVVRNGGFRHARNDVRLVCSFKINDSVT 266



Query Match 99.7%; Score 1532; DB 7; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-156;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFVQOLAHKFTSPQLRMSFIYFS 60  
 DB 27 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFVQOLAHKFTSPQLRMSFIYFS 86  
 QY 61 TRGTTLMKLTEDREQIRQGLEBELQKVLPGGDTVMHEGFERASEQIYYENRQGYRTASVII 120  
 DB 87 TRGTTLMKLTEDREQIRQGLEBELQKVLPGGDTVMHEGFERASEQIYYENRQGYRTASVII 146  
 QY 121 ALTDGELHEDLFFYSEREARNSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 180  
 DB 147 ALTDGELHEDLFFYSEREARNSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 206  
 QY 181 ALQGIHSLILKKSCEILAAEPSTTCAGESFOVVVRNGRFRHARNDVRLCSFKINDSVT 240  
 DB 207 ALQGIHSLILKKSCEILAAEPSTTCAGESFOVVVRNGRFRHARNDVRLCSFKINDSVT 266  
 QY 241 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDG 294  
 DB 267 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDG 320

RESULT 13  
 ADM64576  
 ID ADM64576 standard; protein; 551 AA.  
 AC ADM64576;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Mouse TANGO197-immunoglobulin (Ig) fusion protein.  
 XX  
 KW antibacterial; gene therapy;  
 KW von Willebrand factor A-like domain amino acid sequence;  
 KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
 KW inhalation anthrax; mouse; TANGO197; immunoglobulin; Ig; fusion protein;  
 KW gene.  
 XX  
 OS Mus sp.  
 OS Synthetic.  
 XX  
 FN US2003134786-A1.  
 XX  
 PD 17-JUL-2003.  
 XX  
 PF 20-DEC-2001; 2001US-00038307.  
 XX  
 PR 20-DEC-2001; 2001US-00038307.  
 XX  
 PA (ROTT/) ROTTMAN J B.  
 PA (OKEE/) O'KEEFE T L.  
 PA (OZKA/) OZKAYNAK E.  
 PA (HEAL/) HEALEY J J.  
 XX  
 PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
 XX  
 DR WPI; 2003-829643/77.  
 DR N-PSDB; ADM64575.  
 XX  
 PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
 PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
 PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
 PT to the vWF.  
 XX  
 PS Claim 44; SEQ ID NO 18; 64pp; English.  
 XX  
 CC The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the vWF. Also described are: a method of preventing or

CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to Bacillus anthracis;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
 CC sequence of a fusion protein comprising mouse TANGO197 and immunoglobulin  
 CC (Ig) that can be used to treat exposure to or prevent a symptom of  
 CC anthrax.  
 CC  
 XX Sequence 551 AA;  
 XX

Query Match 99.7%; Score 1532; DB 7; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-156;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFVQOLAHKFTSPQLRMSFIYFS 60  
 DB 27 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFVQOLAHKFTSPQLRMSFIYFS 86  
 QY 61 TRGTTLMKLTEDREQIRQGLEBELQKVLPGGDTVMHEGFERASEQIYYENRQGYRTASVII 120  
 DB 87 TRGTTLMKLTEDREQIRQGLEBELQKVLPGGDTVMHEGFERASEQIYYENRQGYRTASVII 146  
 QY 121 ALTDGELHEDLFFYSEREARNSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 180  
 DB 147 ALTDGELHEDLFFYSEREARNSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 206  
 QY 181 ALQGIHSLILKKSCEILAAEPSTTCAGESFOVVVRNGRFRHARNDVRLCSFKINDSVT 240  
 DB 207 ALQGIHSLILKKSCEILAAEPSTTCAGESFOVVVRNGRFRHARNDVRLCSFKINDSVT 266  
 QY 241 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDG 294  
 DB 267 LNEKPFSEVDYLLCPAPILKEVGKKAALQVSNMDGLSFISSVIITTHCSDG 320

RESULT 14  
 AAE01469  
 ID AAE01469 standard; protein; 403 AA.  
 AC AAE01469;  
 XX  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE Human gene 4 encoded secreted protein HMLFR02, SEQ ID NO:125.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification;  
 KW chromosome 19.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FH Peptide 1..27  
 FT /label= signal\_peptide  
 FT Protein 28..403  
 FT /note= "Mature human secreted protein"  
 FT Misc-difference 175  
 FT /label= Unknown  
 FT /note= "Encoded by GKT"  
 FT Misc-difference 320  
 FT /label= Unknown  
 FT /note= "Encoded by SGT"  
 FT Misc-difference 331  
 FT /label= Unknown





```

Qy 1 GGGRRRDEGSPACYGFDLYFIIDKSGSVLHMHNEIYFVEQLAHKFI SPOLRMSFIYFS 60
   |||||
Db 27 GGGRRRDEGSPACYGFDLYFIIDKSGSVLHMHNEIYFVEQLAHKFI SPOLRMSFIYFS 86
   |||||
Qy 61 TRGTTLMKLTEDREQIRQGLEBELQVLPBGDPTVMHGFERASEQIYYENRQGYRTASVTI 120
   |||||
Db 87 TRGTTLMKLTEDREQIRQGLEBELQVLPBGDPTVMHGFERASEQIYYENRQGYRTASVTI 146
   |||||
Qy 121 ALTDGELHEDLFFYSEREANRSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 180
   |||||
Db 147 ALTDGELHEDLFFYSEREANRSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 206
   |||||
Qy 181 ALQGIHSHILKKSCEIILAEBSSTICAGESFOVVVRGNGFRHANNVDRVLCSPKINDSVT 240
   |||||
Db 207 ALQGIHSHILKKSCEIILAEBSSTICAGESFOVVVRGNGFRHANNVDRVLCSPKINDSVT 266
   |||||
Qy 241 LNEKPFSEVDITYLCPAPILKEVGKALQVSMNDGLSFISSVIITTHCSDGS 295
   |||||
Db 267 LNEKPFSEVDITYLCPAPILKEVGKALQVSMNDGLSFISSVIITTHCSDXS 321
   |||||

```

Search completed: June 13, 2005, 19:56:46  
 Job time : 111.221 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 13, 2005, 19:37:47 ; Search time 99.0301 Seconds  
(without alignments)  
1525,429 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321  
Perfect score: 1536  
Sequence: 1 GGGGRREDGGPACYGDFDY.....GLSPSSSVITTTTCSDGS 295

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536	100.0	564	1 ATRI_HUMAN	Q956X2 homo sapien
2	1519	98.9	562	1 ATRI_MOUSE	Q95252 mus musculu
3	790.5	51.5	489	1 ATR2_HUMAN	P58335 homo sapien
4	770	50.1	487	2 G6DFX2	G6DFX2 mus musculu
5	606	39.5	641	2 O8BVM2	O8BVM2 mus musculu
6	388	25.3	97	2 Q96EC6	Q96EC6 homo sapien
7	159.5	10.4	1161	1 ITAD_RAT	Q95677 ratius norv
8	156	10.2	1162	1 ITAD_HUMAN	Q13349 homo sapien
9	153.5	10.0	1332	2 Q9BPG8	Q9BPG8 halocynthia
10	148.5	9.7	1626	2 O8NPM1	O8NPM1 homo sapien
11	148	9.6	441	2 O8TSC3	O8TSC3 mytilus edu
12	148	9.6	444	2 O8TSC5	O8TSC5 mytilus edu
13	145.5	9.5	1152	1 ITAM_HUMAN	P11215 homo sapien
14	145	9.4	453	2 O8T5C2	O8T5C2 mytilus gal
15	143.5	9.3	1163	1 ITAX_HUMAN	P20702 homo sapien
16	143	9.3	3124	1 CA1C_CHICK	P13944 gallus gall
17	141	9.2	1153	1 ITAM_MOUSE	P05555 mus musculu
18	139	9.0	724	2 Q04588	Q04588 elmeria max
19	139	9.0	3119	1 CA1C_MOUSE	Q06848 mus musculu
20	134.5	8.8	712	2 O43981	O43981 elmeria ten
21	134	8.7	760	2 O70350	O70350 mus musculu
22	132.5	8.6	765	2 Q9U8J9	Q9U8J9 neospora ca
23	131.5	8.6	517	2 Q03853	Q03853 homo sapien
24	131	8.5	760	1 C02_MOUSE	P21180 mus musculu
25	131	8.5	3063	1 CA1C_HUMAN	Q95715 homo sapien
26	130.5	8.5	920	2 Q28984	Q28984 sus scrofa
27	130	8.4	1888	1 CA1E_CHICK	P32018 gallus gall
28	129.5	8.4	1182	2 O8C6T9	O8C6T9 mus musculu
29	128	8.3	221	2 Q99L17	Q99L17 mus musculu
30	128	8.3	3183	2 Q65ZC2	Q65ZC2 caenorhabdi
31	128	8.3	3767	1 M0A3_CAEEL	P34576 caenorhabdi

32	123	8.0	1151	2 Q9J130	Q9J130 ratius norv
33	121	7.9	758	2 Q6W373	Q6W373 ratius norv
34	121	7.9	758	2 Q8CIP8	Q8CIP8 ratius norv
35	121	7.9	2401	2 Q7RF52	Q7RF52 plasmodium
36	120.5	7.8	790	2 Q6DC06	Q6DC06 xenopus lae
37	118.5	7.7	440	2 O8C8T2	O8C8T2 mus musculu
38	118.5	7.7	848	2 O8C720	O8C720 mus musculu
39	118.5	7.7	919	2 Q75R52	Q75R52 lymanaea sta
40	118.5	7.7	3567	2 Q9ES77	Q9ES77 mus musculu
41	118	7.7	652	2 Q9SLI2	Q9SLI2 bos taurus
42	118	7.7	929	1 CA1C_NORVI	Q91145 notophthalm
43	117.5	7.6	637	2 O8IVY1	O8IVY1 homo sapien
44	117.5	7.6	1284	2 Q6P159	Q6P159 homo sapien
45	117.5	7.6	1329	1 KFI0_HUMAN	Q9P218 homo sapien

## ALIGNMENTS

RESULT 1  
ID ATRI\_HUMAN STANDARD; PRT; 564 AA.  
AC Q956X2: 096P02: Q9NVP3:  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 05-JUN-2004 (Rel. 44, Last annotation update)  
DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
GN Name=ANTXR1; Synonyms=ATR, TEM8;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20407466; PubMed=10947988; DOI=10.1126/science.289.5482.1197;  
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.B.,  
RA Montgomery E., Lai A., Riggs G.J., Lengauer C., Vogelstein B.,  
RA Kinzler K.W.;  
RT "Gene expressed in human tumor endothelium";  
RL Science 289:1197-1202(2000)..  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH ANTHRAX TOXIN.  
RX MEDLINE=21557240; PubMed=11700562; DOI=10.1038/n5101998;  
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;  
RT "Identification of the cellular receptor for anthrax toxin";  
RL Nature 414:225-229(2001).  
RN [3]  
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.  
RP (ISOFORM 3).  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Ohtsuki T., Sugiyama T., Irie R.,  
RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine N., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Iizawa T., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagashima K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Nishimura K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,  
RA Ieshima S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta Y.,  
RA Kuwano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
RA Imose N., Mutsaers K., Yuuki H., Ohshima A., Sasaki N., Aotsuka S.,  
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Motiya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Hishigawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujizara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe Y., Komatsu T.,  
RA Mitsuhashi-Sugano J., Satoh T., Shitara Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuko Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RT Nat. Genet. 36:40-45 (2004).  
RN [4]  
RN SEQUENCE FROM N.A. (ISOFORM 4).  
RP TISSUE=Kidney;  
RC MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RX Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Ditschenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bockak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield Y.S.N., Krzywinski M.I., Skalski U., Smallus D.E.,  
RA Schercher A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [5]  
RN INTERACTION WITH ANTHRAX TOXIN.  
RP TISSUE=Placenta;  
RC MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
RX Scoble H.M., Ratney G.U.A., Bradley K.A., Young J.R.T.,  
RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
RT receptor.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174 (2003).  
RN [6]  
RN SPLICING ISOFORM (S) THAT ARE POTENTIAL NMD TARGET (S).  
RP PubMed=14759288; DOI=10.1186/gb-2004-5-2-r8;  
RX Hillman R.T., Green R.B., Brenner S.B.,  
RA "An unappreciated role for RNA surveillance.";  
RT Genome Biol. 5:RESEARCH008.1-RESEARCH008.16 (2004).  
CC -1- FUNCTION: Cellular role is not yet known.  
CC -1- SUBUNIT: Binds to the protective antigen (PA) of *Bacillus*  
CC anthracis. Binding does not occur in the presence of calcium.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Experimental confirmation may be lacking for some  
CC isoforms;  
CC Name=1;  
CC IsoId=Q9H6X2-1; Sequence=Displayed;  
CC Note=May be produced at very low levels due to a premature stop  
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
CC Name=2;  
CC IsoId=Q9H6X2-2; Sequence=VSP\_000444, VSP\_000445;  
CC Name=3;  
CC IsoId=Q9H6X2-3; Sequence=VSP\_000446, VSP\_000447;  
CC Name=4;  
CC IsoId=Q9H6X2-4; Sequence=VSP\_000448, VSP\_000449;  
CC -1- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells  
CC but not in normal endothelial cells  
CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain.  
CC -1- SIMILARITY: Belongs to the ATR family.  
CC -1- SIMILARITY: Contains 1 VWA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF279145; AAK52094.1; -  
DR EMBL: AF421380; AAL26496.1; -  
DR EMBL: AK025429; BAB15128.1; ALT\_INIT.  
DR EMBL: AK001463; BAA91707.1; ALT\_FRAME.  
DR EMBL: BC012074; AAI12074.1; -  
DR HGNC: HGNC:21014; ANTXRL.  
DR H-InvDB: HIX0002125; -  
DR MIM: 606410; -  
DR InterPro: IPR008400; Anth\_Ig.  
DR InterPro: IPR008399; Ant\_C.  
DR InterPro: IPR020315; VWF\_A.  
DR Pfam: PF05587; Anth\_Ig; 1.  
DR Pfam: PF05586; Ant\_C; 1.  
DR Pfam: PF00092; VWA; 1.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS50234; VWF; 1.  
KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 32  
FT CHAIN 1 564  
FT DOMAIN 33 321  
FT TRANSMEM 322 342  
FT DOMAIN 343 564  
FT DOMAIN 44 215  
FT DOMAIN 360 368  
FT DOMAIN 506 564  
FT CARBOHYD 166 166  
FT CARBOHYD 184 184  
FT CARBOHYD 262 262  
FT VARSPPLIC 365 368  
FT VARSPPLIC 369 564  
FT VARSPPLIC 268 297  
FT VARSPPLIC 298 564  
FT VARSPPLIC 319 333  
FT VARSPPLIC 334 564  
SQ SEQUENCE 564 AA; 62789 MW; B18A00AD5DF2233 CRC64;  
Query Match 100.0%; Score 1536; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2,4e-114;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGRRRBDGGPACGPGDLYFLLDKSGSVLAHNNETYYFVQLAKRIFSPQLRMSFIVFS 60  
DB 27 GGGRRRBDGGPACGPGDLYFLLDKSGSVLAHNNETYYFVQLAKRIFSPQLRMSFIVFS 86  
QY TRGTTLMKLTEDBEQIQGLEIQLKVPGLGDTYVHGEFERRASQIYYENRQGYRTASVII 120  
DB TRGTTLMKLTEDBEQIQGLEIQLKVPGLGDTYVHGEFERRASQIYYENRQGYRTASVII 146  
QY 121 ALTDGSLHEDLFFYSEREARNSRDLAGIYVCVGKDFNEFQLARIADSKDHVPVNDGFO 180  
DB 147 ALTDGSLHEDLFFYSEREARNSRDLAGIYVCVGKDFNEFQLARIADSKDHVPVNDGFO 206  
QY 181 ALGGIHSILKSGCETILAEPSITCGSEFOVVVRNGRHRANRVLCSPFINDSVT 240  
DB 207 ALGGIHSILKSGCETILAEPSITCGSEFOVVVRNGRHRANRVLCSPFINDSVT 266  
QY 241 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMNGLSFISSSVYITTHCSDS 295  
DB 267 LNEKPSVEDTYLLCPAPILKEVGKKAALQVSNMNGLSFISSSVYITTHCSDS 321

```

RESULT 2
AC ATR1_MOUSE STANDARD; PRT; 562 AA.
ID ATR1_MOUSE
AC OGC252;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).
GN Name=Atxrl; Synonyms=Atlr; Tem8;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21443268; PubMed=11559528;
RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,
RA Kinzler K.W., St Croix B.;
RT "Cell surface tumor endothelial markers are conserved in mice and
RL humans."
RN [2]
RP SEQUENCE OF 98-562 FROM N.A. (ISOFORM 2).
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikielido I., Oseco N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Busic V., Chochia C., Corbani L.E., Cousins S.,
RA Dalia E., Dregani T.A., Fletcher C.F., Forrest A., Frezer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grilmond S., Guerinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numa K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Piliat R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Borls A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carinini P., Hayatsu N.,
RA Hirczane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Atakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imocani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
CC -1- FUNCTION: Cellular role is not yet known.
CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus
CC anthracis (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9CZ52-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9CZ52-2; Sequence=VSP_000450;
CC Note=No experimental confirmation available;
CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain
CC (By similarity).
CC -1- SIMILARITY: Belongs to the ATR family.
CC -1- SIMILARITY: Contains 1 VWA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF378762; AAL11999.1; -.
DR EMBL; AK013005; BAB28591.1; ALT_INIT.
DR MGD; MGI:1916788; Atxrl.
DR InterPro; IPR008400; Anth_19.
DR InterPro; IPR008399; Ant_C.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF05587; Anth_19; I.
DR Pfam; PF05586; Ant_C; 1.
DR Pfam; PF00092; VWA; 1.
DR PROSITE; PS0234; VWA; 1.
KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.
FT SIGNAL 1 30
FT CHAIN 1 562
FT DOMAIN 31 319 Extracellular (Potential).
FT TRANSMEM 320 340 Potential.
FT DOMAIN 341 562 Cytoplasmic (Potential).
FT DOMAIN 42 213 VWA.
FT DOMAIN 358 366 Asp/Glu-rich (highly acidic).
FT DOMAIN 501 562 Pro-rich.
FT CARBOHYD 164 164 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 182 182 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 260 260 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 477 562
FT FT
FT FT
FT FT
SQ SEQUENCE 562 AA; 62308 MM; 6AC9204954BB4F7C CRC64;
Query Match 98.9%; Score 1519; DB 1; Length 562;
Best Local Similarity 98.6%; Pred. No. 5.5e-113;
Matches 291; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 GGGRRREDGACGCGFDLYETLDKSGSVLHNMNIYFYFVQLAHKFPISQLRMSFYFS 60
DB 25 GGGRRREDGACGCGFDLYETLDKSGSVLHNMNIYFYFVQLAHKFPISQLRMSFYFS 84
OY 61 TRGTTLTKLTEDRQIQGLLEOKVLPGGDTYVHBEFPERASEQIYYENRGYRTASYII 120
DB 85 TRGTTLTKLTEDRQIQGLLEOKVLPGGDTYVHBEFPERASEQIYYENRGYRTASYII 144
OY 121 ALTDGELEHEDLPFYSEBPNARSRLGAIYVCVGYKDNFTQLARIADSKDHVPVNDGFO 180
DB 145 ALTDGELEHEDLPFYSEBPNARSRLGAIYVCVGYKDNFTQLARIADSKDHVPVNDGFO 204
OY 181 ALQGIHSILKSGSCITELIAPSTICGSEPOVVVRNGRPHANVRVYCSFKINDSVT 240
DB 205 ALQGIHSILKSGSCITELIAPSTICGSEPOVVVRNGRPHANVRVYCSFKINDSVT 264
OY 241 LNEKPFVEVDYLLCPAPILKEVGMKAALQVSNMDGSLFSSSVYITTTCHSDGS 295
DB 265 LNEKPFVEVDYLLCPAPILKEVGMKAALQVSNMDGSLFSSSVYITTTCHSDGS 319
RESULT 3
AC ATR2_HUMAN STANDARD; PRT; 469 AA.
ID ATR2_HUMAN
AC P58335; O86U11; O8NB13; Q96NC7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Anthrax toxin receptor 2 precursor (Capillary morphogenesis protein-2)
DE (CMG-2).
GN Name=ATXR2; Synonyms=CMG2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

OX NCBI\_TaxId=9606;  
 [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=21539596; PubMed=11693410;  
 RA Bell S.E., Mavila A., Salazar R., Bayless K.J., Kanagala S.,  
 RA Maxwell S.A., Davis G.E.;  
 RT "Differential gene expression during capillary morphogenesis in 3D  
 RT collagen matrices: regulated expression of genes involved in basement  
 RT membrane matrix assembly, cell cycle progression, cellular  
 RT differentiation and G-protein signaling.";  
 RL J. Cell Sci. 114:2755-2773(2001).  
 [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH ANTHRAX TOXIN.  
 RP TISSUE=Placenta;  
 RC MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
 RX Scobie H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.;  
 RA "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
 RT receptor.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
 [3]  
 RN SEQUENCE OF 78-489 FROM N.A. (ISOFORM 3), AND SEQUENCE FROM N.A.  
 RP (ISOFORM 4).  
 RC TISSUE=Synovial cell;  
 RX PubMed=14702039; DOI=10.1036/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Tanaka K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahori A., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Puruya T., Kikawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Nimomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiroka S., Chiba Y.,  
 RA Ishida S., Oho Y., Takiguchi S., Watanabe S., Yoshida M., Houcha T.,  
 RA Kusano Y., Taniguchi K., Takahashi T., Fujii A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Komai F., Hara R., Takouchi K., Arita M.,  
 RA Imobe N., Matsuino K., Yuki H., Oshima A., Sasaki N., Aoteuka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kunagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Oktani R., Kawakami T., Noguchi N., Itoh T., Shigeta K., Senba T.,  
 RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 CC -i- FUNCTION: Cellular role is not yet known.  
 CC -i- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 CC anthracis in a divalent cation-dependent manner, with the  
 CC following preference: calcium > manganese > magnesium > zinc.  
 CC Seems to bind to collagen type IV and laminin.  
 CC -i- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1 and 2).  
 CC Secreted (isoform 3). Isoform 1 is expressed at the cell surface  
 CC while isoform 2 is predominantly expressed within the endoplasmic  
 CC reticulum and not at the plasma membrane.  
 CC -i- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=P58335-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P58335-2; Sequence=VSP\_008343;  
 CC Name=3;  
 CC IsoId=P58335-3; Sequence=VSP\_008344, VSP\_008345;  
 CC Note=No experimental confirmation available;

CC	Name=4;
CC	Id=ID=P58335-4; Sequence=VSP_008346;
CC	Note=No experimental confirmation available;
CC	-1-TISSUE SPECIFICITY: Expressed in colon, heart, kidney, lung,
CC	liver, peripheral blood leukocytes, placenta, skeletal muscle,
CC	small intestine and spleen.
CC	-1-DOMAIN: Binding to PA seems to be effected through the VWA domain.
CC	-1-SIMILARITY: Belongs to the ATR family.
CC	-1-SIMILARITY: Contains 1 VWFA domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	at the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to licenses@isb-sib.ch).
CC	-----
DR	EMBL; AY040326; AAA77222.1; -
DR	EMBL; AY233452; AAP04016.1; -
DR	EMBL; AK055636; BAB70976.1; ALT_INIT.
DR	EMBL; AK091721; BAC03731.1; -
DR	GeneW; HGNC:21732; ANTXR2.
DR	MIM; 608041; -
DR	InterPro; IPR008399; Ant_C.
DR	InterPro; IPR02035; VWF_A.
DR	Pfam; PF00586; Ant_C; 1.
DR	Pfam; PF00092; VWA; 1.
KW	PROSITE; PS50234; VWFA; 1.
DW	Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.
FT	SIGNAL 1 33
FT	CHAIN 34 489
FT	DOMAIN 34 318
FT	TRANSMEM 319 341
FT	DOMAIN 342 489
FT	DOMAIN 44 213
FT	CARBOND 250 250
FT	CARBOND 260 260
FT	VASPLIC 213 315
FT	VASPLIC 290 322
FT	VTQAGVKWNHDTHCTFGSGSDPPTSSA (in isoform 3).
FT	/FtId=VSP_008344.
FT	MISSING (In Isoform 3).
FT	/FtId=VSP_008345.
FT	VCIMECKEKLTA -> GRGINFSRVPSQ (in isoform 4).
FT	/FtId=VSP_008346.
FT	P -> A (in Ref. 3; BAC03731).
FT	B9F679DB75B662B7 CRC64,
SEQ	SEQUENCE 489 AA: 53692 MW: 89679 DB:55662B7 CRC64,
QY	Query Match 51.5%; Score 790.5; DB 1; Length 489;
QY	Best Local Similarity 53.2%; Pred.No. 8.9e-55;
QY	Matches 157; Conservative 54; Mismatches 81; Indels 3; Gaps 2
Db	1 GCGGRRDDGGPACGCFDLYFIILDKSGSVYHNMEIYYPVQLAHKFISPOLBMSPIVF 59
Db	26 GGGLLRQEQBPSCRRADLYFLVDKSSVANMTEIYNFVQLERFYSPENRLSFIYF 85
QY	60 STGTITLAKLTEDREHRIQGHELOKVPDGDVTYNNHGEPFERASEOIIYYENROGYRTASYI 119
Db	86 SSQAATIFLPFDGRKGISKIGLEDLKRAVPVEBTYIHGELKLANEGI--QKAGGLMTSSII 143
QY	120 IALTDGELEHDLFFTSSEARNSRDLAIYVCYKYKDNEHTOLAILADSKDHVPFPNDGF 179
Db	144 IALTDEKLDGVLPYSAEKEAKISRSLGASVYCVELDFEQALERIADSKREVFPVKGGF 203
QY	180 QALOGIHSILKSKSCITELAAEPSITGCESFOVVVRNGGFPHANVDRILCSFKINDSY 239
Db	204 QALGIINSILAOSTETTELLEIQPSSVCYGEFEQIYLSRGFMILSGSRNSVLCITYNNEY 263
QY	240 TLNEPFSEVEDTYLLCAPRIKEVGMAKALOVSMDGLSFSSIVIIITTHCSDG 294

Db 264 TTSVAPSVQINSMCLCPAPILINKAGETLDSVSPFGKGSVTSGLTIVATCSNG 318

RESULT 4

06DPX2 PRELIMINARY; PRT; 487 AA.

AC Q6DPX2; 25-OCT-2004 (TREMblrel. 28, Created)

DT 25-OCT-2004 (TREMblrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMblrel. 28, Last annotation update)

DE Anthrax toxin receptor 2.

GN Name=Anthr2;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RA Klausegger R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marzella K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boeck S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting R.M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Skolnick J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinska M.I., Skolnick U., Smalick D.E., Scherch A., Schein J.B., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6; TISSUE=Eye;

RC Strainberg R.;

RA Submitted (JUL--2004) to the EMBL/GenBank/DBJ databases.

RL EMBL: BC076595; AAF76595.1; --

DR GO: 0016021; C: integral to membrane; IEA.

DR GO: 0004872; F: receptor activity; IEA.

DR InterPro: IPR008400; Ant\_IG.

DR InterPro: IPR008399; Ant\_G.

DR InterPro: IPR02035; VWF\_A.

DR Pfam: PF05587; Anth\_IG1.

DR Pfam: PF05586; Ant\_G1.

DR SMART: SM00327; VWA; 1.

DR PROSITE: PS50234; VWF\_A; 1.

KM Receptor.

SO SEQUENCE 487 AA; 53184 MM; 61A400D60BCD869 CRC64;

Query Match 50.1%; Score 770; DB 2; Length 487;

Best Local Similarity 52.8%; Pred. No. 3.9e-53;

Matches 150; Conservative 52; Mismatches 80; Indels 2; Gaps 1;

Qy 11 PACGGGDIYPIIDKSSVLMHMEIYYPVQGLAHKRTISPOLRMSFVPSFGRTTLMLKT 70

Db 37 PSCKKADLVYFDKSSGVANWMIETYNFVQLTERFVSPMRISFTVSSQATIIILPT 96

Qy 71 EDREIQIGLEELQKVPFGDPTVHGSEFERSQIYVENRGYRTASVITALTGGEIHD 130

Db 97 GDRKIKIGLELDKAVKVPVGETTYHKGKLANEQI--QNGAGLKASSIITLITGKIDGL 154

Qy 131 LFFYSREANRSDLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFOALQGIHSIL 190

Db 155 VPSYAENBAKKSRLGASVYCVGVLDFEQALERIASKQVFPVKGFOALKGIHSIL 214

Qy 191 KKSCTEILAAEPSTICGSEFOVVVNGFRHANVDRVYCSFINDSVTLNKPFSVD 250

Db 215 AQSTTEILESPSSVCGEFPQVLTTRAVTYSISHDSGLCTFTANSTYTKSEKVSIO 274

Qy 251 TYLLCPAPILKEVGMKAALQVSMNDGLSFISSVYITTTTCSDG 294

Db 275 SSILCPAPVINKGSETLEVSISYDGSASRSRLTIRATECTNG 318

RESULT 5

06BVM2 PRELIMINARY; PRT; 641 AA.

AC 06BVM2; 01-MAR-2003 (TREMblrel. 23, Created)

DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)

DE Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493443011 product:hypothetical Prolin-rich region/von Willebrand factor type A domain containing protein, full insert sequence.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

RA Klausegger R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marzella K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boeck S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting R.M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Skolnick J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinska M.I., Skolnick U., Smalick D.E., Scherch A., Schein J.B., Jones S.J., Marra M.A.; "High-efficiency full-length cDNA cloning.";

RT Meth. Enzymol. 303:19-44(1999).

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=99279253; Pubmed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

RL [3]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Testis;

RC The FANTOM Consortium.

RA The RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

RL [4]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=20499374; Pubmed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";

RL Genome Res. 10:1617-1630(2000).

RL [5]

RN SEQUENCE FROM N.A.

RP STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=20530913; Pubmed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoke S., Sasaki N., Carninci P., Kono H., Akiyama U., Nishi K., Kitanai T., Tashiro H., Itoh M., Sumi M., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";

RL Genome Res. 10:1757-1771(2000).

RL [6]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA Adachi J., Aizawa K., Akimura T., Arikawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hasehima W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraka T., Hirose T.,  
 RA Hoti F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akehira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK077206; BAC36683.1; -  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR006400; Arch\_19.  
 DR InterPro; IPR02035; VMP\_A.  
 DR Pfam; PF05587; Antch\_19; 1.  
 DR Pfam; PF00092; VMA; 1.  
 DR PRINTS; PRO0453; VMPADOMAIN.  
 DR SMART; SM00327; VMA; 1.  
 DR PROSITE; PS50234; VMPA; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 641 AA; 70415 MW; 199E300730BC85E3 CRC64;

Query Match 39.5%; Score 606; DB 2; Length 641;  
 Best Local Similarity 42.2%; Pred. No. 7,2e-40;  
 Matches 119; Conservative 58; Mismatches 105; Indels 0; Gaps 0;

QY 10 GPACGCGFDLYFIIDKSGSVLHNNELIYFVQALAKHFIQPMSPFVSTGTTMLKL 69  
 DB 68 GDDCGGIFDLIVLDKSGSVADNWIHYSFAGELVKFTFNHNLISITTSYEAELIPL 127  
 QY 70 TEDREQIROGLEELQKVLPGSDTYMHGFEFASFOIYENNGQRTASVILALDGEIHE 129  
 DB 128 TSDSEKINKSLVLKSVIPQGLTHMQGLRANQIRKSTLGGRIVNSVIALTDGLLL 187  
 QY 130 DLFFYSEREANRSHDLGAIYVCVGVKQFNETOLARLADSKDHPVPVNDGFOALGIIHSI 189  
 DB 188 KPYLDITMEAKKARRMGAIVTVGVFMYSKQGLVINAGDPRCFVGEGFALGAVDPL 247  
 QY 190 LKKSCEIILAAPESTTCAGSFQVVVRNGFRHARNDRVLCSPKINDSVTLNKPSPSVE 249  
 DB 248 TSKECTEILSVQPTVYVCAKDFYQVNIISGHGLNNTSNMKQVLCRFKPSDKVDESPDMN 307  
 QY 250 DTVLLCPAPILKEVGMKALGVSMNDGLSFSSSVIITTHC 291  
 DB 308 EHSITCPGPKIKHTGEDVSLQVSLNGLISFGNKLITSTNC 349

## RESULT 6

Q96EC6 PRELIMINARY; PRT; 97 AA.  
 AC Q96EC6 01-DEC-2001 (TREMBlrel. 19; Created)  
 DT 01-DEC-2001 (TREMBlrel. 19; Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25; Last annotation update)  
 DE ANTXR1 protein (Fragment).  
 GN Name=ANTXR1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast;  
 RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Straube R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,  
 RA Alschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Tohyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.D., Abremson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez M.I., Skalka U., Smutiz J., Myers R.M., Butterfield Y.S.,  
 RA Krzyzanski M.I., Skalka U., Smutiz J., Myers R.M., Butterfield Y.S.,  
 RA Jones S.J., Maiz M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast;  
 RA Straube R.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012475; AAH12475.1; -  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR006400; Antch\_19.  
 DR Pfam; PF05587; Antch\_19; 1.  
 DR NON TER 1  
 PT 1  
 SQ SEQUENCE 97 AA; 10453 MW; 14F475F0B170E71A CRC64;

Query Match 25.3%; Score 388; DB 2; Length 97;  
 Best Local Similarity 95.1%; Pred. No. 1.9e-23;  
 Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 211 FQVVGNGFRHARNDRVLCSPKINDSVTLNKPSPSVEVDTYLLCPAPILKEVGMKALQ 270  
 DB 1 FQVVGNGFRHARNDRVLCSPKINDSVTLNKPSPSVEVDTYLLCPAPILKEVGMKALQ 60  
 QY 271 VSMNDGLSFSSSVIITTHCS 292  
 DB 61 VSMNDGLSFSSSVIITTHCS 82

## RESULT 7

ITAD\_RAT STANDARD; PRT; 1161 AA.  
 AC OJOYE7;  
 DT 05-JUL-2004 (Rel. 44; Created)  
 DT 05-JUL-2004 (Rel. 44; Last sequence update)  
 DT 05-JUL-2004 (Rel. 44; Last annotation update)  
 DE Integrin alpha-D precursor.  
 GN Name=Itgad;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA O'Brien M.M., VanderVlieten M., Kilgannon P.D., Dietrich G.,  
 RA Gallatin W.M.;  
 RT "Cloning of rat alpha D, a novel beta 2 integrin.";  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and  
 CC VCAM1. May play a role in the atherosclerotic process such as  
 CC clearing lipoproteins from plaques and in phagocytosis of blood-  
 CC borne pathogens, particulate matter, and senescent erythrocytes  
 CC from the blood (By similarity).  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D  
 CC associates with beta-2 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1- DOMAIN: The integrin I-domain (insert) is a VMPA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.



CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 WFPA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, AF021334; AAF21241.1; --  
 CC HSBP, P1215; IBO.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; WFPA.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; WFA.  
 DR PRINTS; PR01185; INTEGRINA.  
 DR PRINTS; PR00453; WFPA DOMAIN.  
 DR SMART; SMO0191; Int\_alpha; 5.  
 DR SMART; SMO0327; WMA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS00234; WFPA; 1.  
 DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
 KW Repeat; Signal; Transmembrane.  
 FT SIGNAL 1 19 Potential.  
 FT CHAIN 20 1161 Integrin alpha-D.  
 FT DOMAIN 20 1100 Extracellular (Potential).  
 FT TRANSMEM 1101 1121 Potential.  
 FT DOMAIN 1122 1161 Cytoplasmic (Potential).  
 FT REPEAT 34 87 FG-GAP 1.  
 FT REPEAT 88 7 FG-GAP 2.  
 FT REPEAT 152 334 WFPA.  
 FT REPEAT 352 402 FG-GAP 3.  
 FT REPEAT 403 454 FG-GAP 4.  
 FT REPEAT 456 517 FG-GAP 5.  
 FT REPEAT 519 577 FG-GAP 6.  
 FT REPEAT 582 634 FG-GAP 7.  
 FT CA\_BIND 467 475 Potential.  
 FT CA\_BIND 531 539 Potential.  
 FT CA\_BIND 594 602 Potential.  
 FT SITE 1126 1130 GEFK motif.  
 FT DISULFID 69 76 By similarity.  
 FT DISULFID 108 126 By similarity.  
 FT DISULFID 656 711 By similarity.  
 FT DISULFID 769 775 By similarity.  
 FT DISULFID 845 860 By similarity.  
 FT DISULFID 993 1017 By similarity.  
 FT DISULFID 1022 1027 By similarity.  
 FT CARBOHYD 61 61 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 89 89 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 245 245 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 393 393 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 696 696 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 734 734 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 784 784 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 907 907 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 936 936 N-linked (GLCNAc... ) (Potential).  
 FT CARBOHYD 1045 1045 N-linked (GLCNAc... ) (Potential).  
 FT SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;  
 Query Match 10.4%; Score 159.5; DB 1; Length 1161;  
 Best Local Similarity 27.9%; Pred. No. 0.00079;  
 Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;  
 QY 18 DLYPLDKSGSV-LHHNWEIYVYEQLAHKPTSPQLMSFTVFTSGTTLNKLTE----- 71  
 DB 152 DIAFLIDSGSINRGDPAQMDPVALKGEFASSTLPSLMQYSNLTKEHFTFEFNKIL 211  
 QY 72 DREQIRQGLEELQKLRPGSDPTMHGEGFRASQIYERKQGYRTA-SYIILTLTGELHED 130  
 DB 212 DPGSLVDPITVQLQ-----GLTYTATGIRTWBELFHSNKGSKAKKILVITDQGYRD 266

QY 131 LFFYSE--REANRSNDLGAIVYCVGKD-FNE-TQLARI-----ADSKDHVFPVNDGFOA 181  
 DB 267 PLEKSDVLPADKA---GIIRYALGVGDAFQEPALKEKLNITGSAFPQDDHFKVGN-FAA 322  
 QY 182 LOGIHSILKSKSCIEIILAEPTTICAGSEFPVVYVNGCFRIARNVD 227  
 DB 323 LRSIORQLOEK-----IFAIEGTQSRSSSSFOHEMSQEGFSALTSID 364  
 RESULT 8  
 ITAD\_HUMAN STANDARD; PRT; 1162 AA.  
 AC Q13349; Q15575; Q15576;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DE Integrin alpha-D precursor (LeukoIntegrin alpha D) (CD11d) (ADB2).  
 GN Name=ITGAD;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=96111956; PubMed=8777714; DOI=10.1016/1074-7613(95)90058-6;  
 RA Van der Vlieten M., Le Trong H., Wood C.L., Moore P.F., St John T.,  
 RT Staunton D.E., Gallatin W.M.;  
 RL "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3."  
 RN Immunity 3:683-690(1995).  
 [2]  
 RP SEQUENCE OF 1-235 FROM N.A.  
 RX MEDLINE=20187620; PubMed=10722744; DOI=10.1074/jbc.275.12.8959;  
 RA Noci J.D., Johnson A.K., Dillon J.D.;  
 RT "Structural and functional characterization of the leukocyte integrin gene CD11d. Essential role of Sp1 and Sp3."  
 RL J. Biol. Chem. 275:8959-8969(2000).  
 [3]  
 RP SEQUENCE OF 467-537, 571-602, 633-666, 788-834 AND 910-1125 FROM N.A.  
 RX MEDLINE=96257236; PubMed=8666289; DOI=10.1016/0378-1119(95)00869-1;  
 RA Wong D.A., Davis E.M., Lebeau M., Springer T.A.;  
 RT "Cloning and chromosomal localization of a novel gene-encoding a human beta 2-integrin alpha subunit."  
 RL Gene 171:291-294(1996).  
 [4]  
 RP INTERACTION WITH VCAM1.  
 RX MEDLINE=99059842; PubMed=9841932;  
 RA Grayson M.H., Van der Vlieten M., Sterbinsky S.A., Michael Gallatin W.,  
 RT Hoffman P.A., Staunton D.E., Bochner B.S.;  
 RL "alpha2beta2 integrin is expressed on human eosinophils and functions as an alternative ligand for vascular cell adhesion molecule 1 (VCAM-1)."  
 RN J. Exp. Med. 188:2187-2191(1998).  
 [5]  
 RP INTERACTION WITH VCAM1.  
 RX MEDLINE=99370002; PubMed=10438935;  
 RA Van der Vlieten M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,  
 RT Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;  
 RL "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a binding interface between I domain and VCAM-1."  
 RN J. Immunol. 163:1984-1990(1999).  
 -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and VCAM1. May play a role in the atherosclerotic process such as clearing lipoproteins from plaques and in phagocytosis of blood-borne pathogens, particulate matter, and senescent erythrocytes from the blood.  
 -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D associates with beta-2.  
 -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 -1- TISSUE SPECIFICITY: Expressed moderately on myelomonocytic cell lines and subsets of peripheral blood leukocytes and strongly on

tissue-specialized cells, including macrophages foam cells within atherosclerotic plaques, and on splenic red pulp macrophages.

-1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.

-1- SIMILARITY: Belongs to the integrin alpha chain family.

-1- SIMILARITY: Contains 7 FG-GAP repeats.

-1- SIMILARITY: Contains 1 VWFA domain.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

-----

DR EMBL: U37028; AAB36547.1; -.

DR EMBL: U40274; AAB60634.1; -.

DR EMBL: U40275; AAB60635.1; -.

DR EMBL: U40276; AAB60636.1; -.

DR EMBL: U40277; AAB60637.1; -.

DR EMBL: U40279; AAB60638.1; -.

DR EMBL: U40278; AAB60638.1; JOINED.

DR EMBL: AF187881; AAF62875.1; -.

DR HSSP: P11215; 1BHQ.

DR Genew: HGNC:6146; ITGAD.

DR MIM: 602453; -.

DR GO: GO:0008305; C:integrin complex; TAS.

DR GO: GO:0016337; P:cell-cell adhesion; NAS.

DR GO: GO:0007160; P:cell-matrix adhesion; NAS.

DR GO: GO:0006955; P:immune response; NAS.

DR InterPro: IPR000413; Integrin\_alpha.

DR InterPro: IPR002035; VWF\_A.

DR Pfam: PF01839; FG-GAP; 3.

DR Pfam: PF00357; Integrin\_alpha; 1.

DR Pfam: PF00092; VWA; 1.

DR PRINTS: PRO1185; INTEGRINA.

DR PRINTS: PRO0453; VWFADOMAIN.

DR SMART: SM00191; Int\_alpha; 5.

DR SMART: SM00327; VWA\_1.

DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.

DR PROSITE: PS50234; VWF\_A; 1.

DR KEGG: Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor; Repeat; Signal; Transmembrane.

FT CHAIN 1 17 Potential.

FT SIGNAL 1 17 Integrin alpha-D.

FT DOMAIN 18 1100 Extracellular (Potential).

FT TRANSMEM 1101 1121 Potential.

FT DOMAIN 1122 1162 Cytoplasmic (Potential).

FT REPEAT 32 85 FG-GAP 1.

FT REPEAT 86 ? FG-GAP 2.

FT DOMAIN 150 332 VWFA.

FT REPEAT 350 400 FG-GAP 3.

FT REPEAT 401 452 FG-GAP 4.

FT REPEAT 454 516 FG-GAP 5.

FT REPEAT 518 576 FG-GAP 6.

FT REPEAT 581 633 FG-GAP 7.

FT CA\_BIND 465 473 Potential.

FT CA\_BIND 530 538 Potential.

FT CA\_BIND 593 601 Potential.

FT SITE 1127 1131 GFPKR motif.

FT DISULFID 67 74 By similarity.

FT DISULFID 106 124 By similarity.

FT DISULFID 655 710 By similarity.

FT DISULFID 769 775 By similarity.

FT DISULFID 846 861 By similarity.

FT DISULFID 994 1018 By similarity.

FT DISULFID 1023 1028 By similarity.

FT CARBOHYD 59 59 N-linked (GlcNAc...)

FT CARBOHYD 87 87 N-linked (GlcNAc...)

FT CARBOHYD 99 99 N-linked (GlcNAc...)

FT CARBOHYD 391 391 N-linked (GlcNAc...)

FT CARBOHYD 691 691 N-linked (GlcNAc...)

FT CARBOHYD 733 733 N-linked (GlcNAc...)

FT CARBOHYD 873 873 N-linked (GlcNAc...)

FT CARBOHYD 957 957 N-linked (GlcNAc...)

FT CARBOHYD 1046 1046 N-linked (GlcNAc...)

FT CONFLICT 500 500 Missing (in Ref. 2)

FT CONFLICT 515 518 GHPW -> ATP (in Ref. 2)

FT CONFLICT 825 825 L -> V (in Ref. 2)

FT CONFLICT 984 984 V -> A (in Ref. 2)

SO SEQUENCE 1162 AA; 126885 MW; P296AIA3545SD77D CRC64;

Query Match 10.2%; Score 156; DB 1; Length 1162;

Best Local Similarity 24.3%; Pred.No. 0.0015;

Matches 67; Conservative 50; Mismatches 119; Indels 40; Gaps 11;

QY 8 DGGPAC-YGGFDLYFLIDKSGSV-LHHWNIYFVEQLAKFTSPQLRMSPIVSTGTT 65

DB 139 DATPECHQEMDVFLLIDSGSIDQNDPDMKQFVQAVMGFGSTDTLFLMYS-----N 134

QY 66 LMKLTEDREQIRQGLEBLQKLP-----GDTYMHGEFERASEQIYYENRCGYTA-SVIT 120

DB 195 LKIHFTFTQFRTSPSQSLVDPIVOLKGLTFATGILTVVTLFHHKNGARKSAKKILI 254

QY 121 ALTDGELHEDLFYFSEANRSRDLGAIYCVGVKDFENETQLAR-----IADSKDHF 173

DB 255 VITDQKXKPLRYSD-VIPQAKKAGIRYALGVGHAFCQPTARQELNTSSAPPODHVF 313

QY 174 PVNDGFQALQGIHSLIKSKSCEIILAAPESTICAGSFQVYVGVNGFRHARNDVRLC-- 231

DB 314 KV-DNFAALOSIQKQLEK-----IYAVEGQSASASSFQHEMGEGESTALTMDGLFLGA 368

QY 232 --SFKINDS-----VTLNKPSFVEDTYL 253

DB 369 VGSFWSGCAFLYPNMSPTFINMSQENVMDRISYL 404

RESULT 9

Q9BP08 PRELIMINARY; PRT; 1332 AA.

ID Q9BP08

AC Q9BP08

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Integrin alpha Htl precursor.

GN Name=HtrGAL;

OS Halocynthia roretzi (Sea squirt).

OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;

OC Stolidobranchia; Pyuridae; Halocynthia.

OX NCBI\_TaxID=7729;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Hemocyte;

RX MEDLINE=21103187; PubMed=11160215;

RA Miyazawa S., Azumi K., Nonaka M.;

RT "Cloning and characterization of integrin alpha subunits from the

RT solitary ascidian, Halocynthia roretzi.";

RL J. Immunol. 166:1710-1715(2001).

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).

CC -1- SIMILARITY: Belongs to the integrin alpha chain family.

DR EMBL: AB048261; BAB21479.1; -.

DR HSSP: P11215; 1BHQ.

DR GO: GO:0016021; C:integral to membrane; IEA.

DR GO: GO:0008305; C:integrin complex; IEA.

DR GO: GO:0005515; F:protein binding; IEA.

DR GO: GO:0007160; P:cell-matrix adhesion; IEA.

DR GO: GO:0007229; P:integrin-mediated signaling pathway; IEA.

DR InterPro: IPR000413; Integrin\_alpha.

DR InterPro: IPR002035; VWF\_A.

DR Pfam: PF00092; VWA; 1.

DR PRINTS: PRO1185; INTEGRINA.

DR PRINTS: PRO0453; VWFADOMAIN.

DR SMART: SM00191; Int\_alpha; 5.

DR SMART: SM00327; VWA\_1.

DR PROSITE: PS50234; VWF\_A; 1.

KW Cell adhesion; Integrin; Signal; Transmembrane.  
 FT SIGNAL 1 30 Potential.  
 FT CHAIN 31 1332 Integrin alpha H1.  
 SQ SEQUENCE 1332 AA; 145851 MW; 0D9108D2B05CFEAE CRC64;  
 Query Match 10.0%; Score 153.5; DB 2; Length 1332;  
 Best Local Similarity 24.1%; Pred. No. 0.0028; Indels 75; Gaps 10;  
 Matches 63; Conservative 42; Mismatches 81;  
 QY 3 GGRREDG-----PACYGFDLYFTLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFI 57  
 DB 188 GNMKESRNTECS--SGVVLFLVDGSSVGKRFVDKQVKNIT----- 232  
 QY 58 VFSTRGTTLMKLTEDREGIQGLELOKVLPG-----GDTYMHGFERAS 102  
 DB 233 -----AKLDIGKEIVRGVGVQSHYVEGSKINKQYITTEISIGFELDNFENAV 283  
 QY 103 EQIYENRQGR-----ASVIALTDGELHEDLPFYSREANRS 142  
 DB 284 DRI---QLOGYTYTGALOKVIRDFPDAYIGNKQVLLLTLDGAKNKLIP--NANRL 338  
 QY 143 RDLGAIYVCVGKFOENETOLARIA---DSKDFVPVNDGFOALQGIHSILKSCIEILA 199  
 DB 339 RNKGIATPAVGVEYDISELKLINSGTSDTRVFTYTD-FGELDSIVKSLQTEIQSFVLE 397  
 QY 200 AEPSTICAGESPQVVGNGF 220  
 DB 398 GKGSAKTAG--YEMHFGENG 416  
 RESULT 10  
 Q8NFM1  
 ID Q8NFM1 PRELIMINARY; PRT; 1626 AA.  
 AC Q8NFM1  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
 DE Alpha 1 type XXII collagen.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RA Koch M., Jin W., Ashworth T., Burgess R.E.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF406780; AA003620.1; -.  
 DR HSSP; P18614; IIMP.  
 DR Genew; HGNC:22989; COL22A1.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0007155; P:cell adhesion; IEA.  
 DR GO; GO:0006817; P:phosphate transport; IEA.  
 DR InterPro; IPR008161; C1g\_helix.  
 DR InterPro; IPR008160; C1g\_helix.  
 DR InterPro; IPR008985; Cona\_like\_1ec\_g1.  
 DR InterPro; IPR003129; TSP\_N.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01391; Collagen; 16.  
 DR Pfam; PF00092; VMA; 1.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR Prodom; PD000007; C1g\_helix; 7.  
 DR SMART; SM00327; VMA; 1.  
 DR PROSITE; PS50234; VMA; 1.  
 KW Collagen.  
 SQ SEQUENCE 1626 AA; 161115 MW; 34C68B3CFD467407 CRC64;  
 Query Match 9.7%; Score 148.5; DB 2; Length 1626;  
 Best Local Similarity 25.4%; Pred. No. 0.0091;  
 Matches 65; Conservative 45; Mismatches 105; Indels 41; Gaps 14;

QY 9 GGPACVY-----GPDLYFTLDKSGSV-LHHMNEIYFVEQLAHKF-ISP-QLRMSFI 57  
 DB 21 GGGGCCQARAGCKSVHDLVFLDTSVSGKEPFEKROWANLVDFEVPDPTRGVV 80  
 QY 58 VFSTRGTTLMK--LTEDREGIQGLELOKVLPGDITYMHGFERASEQIYENRQ-- 112  
 DB 81 RYSDRPTTAFELGPGSQEYKAAARL--AYHGANTGTDALRYTTARSPFPAAGRPR 138  
 QY 113 ---YRTASVIALTDGELHEDLPFYSREANRSRLGAIYVCVGKDFNETOLARIAD-- 167  
 DB 139 DRAK--QVALILTDR-SQDLVLDAAAHHRA---GIRFPAVGSALEKELEIASP 192  
 QY 168 SKDFVPVNDGFOALQGIHSILKSCIEILAEPTTCAGESPQVVGNGFPHANVD 227  
 DB 193 KSAHFVHVSD-FNAIDIRKLRRLCENVLC--PS-----VAEGDRFHTNGT 240  
 QY 228 RVLCSFKINDSVTLNE 243  
 DB 241 KEITGFDLMDLFSVKE 256  
 RESULT 11  
 Q8T6U5  
 ID Q8T6U5 PRELIMINARY; PRT; 441 AA.  
 AC Q8T6U5;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
 DE Proximal thread matrix protein 1 variant a.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Periomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 NCBI\_Taxid=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;  
 RA Sun C., Lucas J.M., Walte J.H.;  
 RT "Collagen-binding matrix proteins from elastomeric extraorganic  
 RT bysael fibers."  
 RL Biomacromolecules 3:1240-1248(2002).  
 DR EMBL; AF414454; AA083537.1; -.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00092; VMA; 2.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00327; VMA; 2.  
 DR PROSITE; PS50234; VMA; 2.  
 KW Matrix protein.  
 SQ SEQUENCE 441 AA; 47543 MW; 881DBD36B891D2B CRC64;  
 Query Match 9.6%; Score 148; DB 2; Length 441;  
 Best Local Similarity 25.5%; Pred. No. 0.002; Indels 26; Gaps 12;  
 Matches 51; Conservative 46; Mismatches 77;  
 QY 13 CYGFGDLYFTLDKSGSV---LHHMNEIYFVEQLAHKF--ISP-LRMSFIESTRGTT 65  
 DB 236 CAGHADIATFVDASSSTINANNPNVYGLMKDFMDIYDRFMTGPDGTQFVAFADATK 295  
 QY 66 LMKLTG--DREGIQGLELOKVLPG--GDTYMHGFERASEQIYENRQ--YRTASV 118  
 DB 296 QFGKDYSSRAEIKGALD--KTPPSITGTAIGDGENARLEV-FENRSGGGEVEQKV 351  
 QY 119 IIALTDGEL--HEDLPFYSREANRSRLGAIYVCVGKDFNETOLARIADSDHVPV 175  
 DB 352 VILITDQNNGHKS---PEHESLAKESVVALGVGTFKSELINIASSEYVF-T 406  
 QY 176 NDGFOALQGIHSILKSC 195  
 DB 407 TSEFNKLSKIMENVVLAQM 426  
 RESULT 12  
 Q8T5C3

ID Q8TSC3 PRELIMINARY; PRT; 444 AA.  
AC Q8TSC3;  
RT 01-JUN-2002 (TRENBLER. 21, Created)  
DT 01-JUN-2002 (TRENBLER. 21, Last sequence update)  
DE 01-JUN-2003 (TRENBLER. 24, Last annotation update)  
OS Mytilus edulis (Blue mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;  
NCBI\_TaxID=6550;  
RN NCBI\_TaxID=6550;  
RP  
SEQUENCE FROM N.A.  
RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;  
RA Sun C., Lucas J.M., Waite J.H.;  
RT "Collagen-binding matrix proteins from elastomeric extraorganic  
RT bresal fibers.";  
RL Biomacromolecules 3:1240-1248(2002).  
DR EMBL; AY053390; AL17973.1;  
DR GO; GO:0005198; Functional molecule activity; IEA.  
DR InterPro; IPR02035; VWF\_A.  
DR Pfam; PF00092; VMA; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VMA; 2.  
DR PROSITE; PS50234; VWF\_A.  
KM Matrix protein.  
SQ SEQUENCE 444 AA; 47815 MW; D2C605347450C931 CRC64;  
  
Query March 9.6%; Score 148; DB 2; Length 444;  
Best Local Similarity 25.5%; Pred. No. 0.002;  
Matches 51; Conservative 46; Mismatches 77; Indels 26; Gaps 12;  
  
QY 13 CYGQFDYFLDKSGSV---LHMNEIYFVEQLAHKF--ISPO-LRMSFVSTGTT 65  
DB 239 CAGHADIAPVDASSINANNPNVGLMKDKIVRFNKTGPDGQPAVVFADATK 298  
QY 66 LMKLTE--DREQIQGLELOKVLPG--GDTYMEGFRASQIYENRQ--YRTASV 118  
DB 299 QFGIKDYSSKAIKGAID--KTPSIIGQTAIDGLENALFV-FPRNGGKEEVQKV 354  
QY 119 IIALTDEGL--HEDLFYSEBRARSRLGAIYVCVG-KDFNETOLARISDKHPEV 175  
DB 355 VILITDGGNNGHKS---PEHSSILRKGGVIVAIIGTGTFKLSELINISSEYVF-T 409  
QY 176 NDGFQALQGIHSILKSCSI 195  
DB 410 TSSFNKLKSKIMENVKLAQM 429  
  
RESULT 13  
ID ITAM HUMAN STANDARD; PRT; 1152 AA.  
AC P11215;  
RT 01-JUL-1989 (Rel. 11, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha  
DE subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1  
DE (Neutrophil adhesion receptor).  
GN Name=ITGAM; Synonyms=CD11B, CR3A;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_TaxID=9606;  
RN NCBI\_TaxID=9606;  
RP  
SEQUENCE FROM N.A.  
RX MEDLINE=86315033; PubMed=2457584;  
RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor  
RT type 3, CD11b) alpha subunit. Cloning, primary structure, and relation  
RT to the Integrin, von Willebrand factor and factor B.";  
RL J. Biol. Chem. 263:12403-12411(1988).  
RN  
RP SEQUENCE FROM N.A.

RX MEDLINE=86190151; PubMed=2833753;  
RA Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
RT "Molecular cloning of the alpha subunit of human and guinea pig  
RT leukocyte adhesion glycoprotein Mo1: chromosomal localization and  
RT homology to the alpha subunits of integrins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
RN  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86257215; PubMed=2454931; DOI=10.1083/jcb.106.6.2153;  
RA Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion  
RT receptor Mo1 (complement receptor type 3).";  
RL J. Cell Biol. 106:2153-2158(1988).  
RN  
[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93123748; PubMed=8419480;  
RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.P., Tenen D.G.;  
RT "Structural analysis of the CD11b gene and phylogenetic analysis of  
RT the alpha-integrin gene family demonstrate remarkable conservation of  
RT genomic organization and suggest early diversification during  
RT evolution.";  
RL J. Immunol. 150:480-490(1993).  
RN  
[5]  
RP SEQUENCE OF 9-1153 FROM N.A.  
RX MEDLINE=89098933; PubMed=2563162;  
RA Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "cDNA sequence for the alpha M subunit of the human neutrophil  
RT adherence receptor indicates homology to integrin alpha subunits.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RN  
[6]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaout M.A.;  
RT "The promoter of the CD11b gene directs myeloid-specific and  
RT developmentally regulated expression.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RN  
[7]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX TISSUE=Blood;  
RA Pahl H.L., Rosmarin A.G., Tenen D.G.;  
RT "Characterization of the myeloid-specific CD11b promoter.";  
RL Blood 79:865-870(1992).  
RN  
[8]  
RP SEQUENCE OF 17-31.  
RX MEDLINE=87076721; PubMed=3539202; DOI=10.1016/0167-4838(86)90037-3;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaout M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mo1: conservation  
RT across species and homology to platelet IIB/IIIA.";  
RL Biochim. Biophys. Acta 874:368-371(1986).  
RN  
[9]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RX MEDLINE=95171458; PubMed=7867070; DOI=10.1016/S0969-2126(01)00271-4;  
RA Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
RT "Crystal structure of the A domain from the alpha subunit of integrin  
RT CR3 (CD11b/CD18).";  
RL Cell 80:631-638(1995).  
RN  
[10]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RX MEDLINE=96363671; PubMed=8747460; DOI=10.1016/S0969-2126(01)00271-4;  
RA Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (I-domain): a pathway for  
RT activation?";  
RL Structure 3:1333-1340(1995).  
RN  
[11]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RX MEDLINE=98363595; PubMed=9687375; DOI=10.1016/S0969-2126(98)00093-8;  
RA Baldwin E.T., Sarver R.W., Bryant G.L. Jr., Curry K.A.,  
RA Batwinke M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,  
RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,  
RA Mutschler V.I., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "Cation binding to the integrin CD11b I domain and activation model

RT assessment." ;  
 RL structure 6:923-935(1998).  
 RN [12]  
 RP 3D-STRUCTURE MODELING OF 17-616.  
 RX MEDLINE=98226734; PubMed=9560195; DOI=10.1073/pnas.95.9.4870;  
 RA Ovig C., Springer T.A.;  
 RT "Experimental support for a beta-propeller domain in integrin alpha-  
 subunit and a calcium binding site on its lower surface." ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
 CC -1- FUNCTION: Integrin alpha-M/beta-2 is implicated in various  
 adhesive interactions of monocytes, macrophages and granulocytes  
 as well as in mediating the uptake of complement-coated particles.  
 It is identical with CR-3, the receptor for the IC3b fragment of  
 the third complement component. It probably recognizes the R-G-D  
 peptide in C3b. Integrin alpha-M/beta-2 is also a receptor for  
 fibrinogen, factor X and ICAM1. It recognizes P1 and P2 peptides  
 of fibrinogen gamma chain.  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-M  
 associates with beta-2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and  
 granulocytes.  
 CC -1- DOMAIN: The integrin I-domain (insert) is a VFPA domain. Integrins  
 with I-domain do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 VFPA domain.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD11b entry;  
 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.emb-ch/announce/  
 or send an email to license@eb-sib.ch).  
 CC -----  
 DR EMBL/ J03925; AAA59544.1; -;  
 DR EMBL/ M18044; AAA59491.1; -;  
 DR EMBL/ J04145; AAA59803.1; -;  
 DR EMBL/ S52227; AAB24821.1; -;  
 DR EMBL/ S52152; AAB24821.1; JOINED.  
 DR EMBL/ S52153; AAB24821.1; JOINED.  
 DR EMBL/ S52154; AAB24821.1; JOINED.  
 DR EMBL/ S52155; AAB24821.1; JOINED.  
 DR EMBL/ S52157; AAB24821.1; JOINED.  
 DR EMBL/ S52159; AAB24821.1; JOINED.  
 DR EMBL/ S52161; AAB24821.1; JOINED.  
 DR EMBL/ S52164; AAB24821.1; JOINED.  
 DR EMBL/ S52165; AAB24821.1; JOINED.  
 DR EMBL/ S52167; AAB24821.1; JOINED.  
 DR EMBL/ S52169; AAB24821.1; JOINED.  
 DR EMBL/ S52170; AAB24821.1; JOINED.  
 DR EMBL/ S52173; AAB24821.1; JOINED.  
 DR EMBL/ S52174; AAB24821.1; JOINED.  
 DR EMBL/ S52180; AAB24821.1; JOINED.  
 DR EMBL/ S52181; AAB24821.1; JOINED.  
 DR EMBL/ S52184; AAB24821.1; JOINED.  
 DR EMBL/ S52189; AAB24821.1; JOINED.  
 DR EMBL/ S52191; AAB24821.1; JOINED.  
 DR EMBL/ S52192; AAB24821.1; JOINED.  
 DR EMBL/ S52203; AAB24821.1; JOINED.  
 DR EMBL/ S52212; AAB24821.1; JOINED.  
 DR EMBL/ S52213; AAB24821.1; JOINED.  
 DR EMBL/ S52216; AAB24821.1; JOINED.  
 DR EMBL/ S52219; AAB24821.1; JOINED.  
 DR EMBL/ S52220; AAB24821.1; JOINED.  
 DR EMBL/ S52221; AAB24821.1; JOINED.  
 DR EMBL/ S52222; AAB24821.1; JOINED.  
 DR EMBL/ S52226; AAB24821.1; JOINED.  
 DR EMBL/ M76724; AAB58410.1; -;  
 DR EMBL/ M84477; AAB51960.1; -;

DR PIR; A31108; RWHUB.  
 DR PDB; 1ABX; Model; @=17-1152.  
 DR PDB; 1BHO; X-ray; 1/2=-.  
 DR PDB; 1BHQ; X-ray; 1/2=-.  
 DR PDB; 1IDN; X-ray; 1/2=-.  
 DR PDB; 1IDO; X-ray; @=140-331.  
 DR PDB; 1JLM; X-ray; @=143-334.  
 DR PDB; 1MLU; X-ray; A=137-331.  
 DR PDB; 1MF7; X-ray; A=144-337.  
 DR PDB; 1N92; X-ray; A=140-335.  
 DR PDB; 1NA5; X-ray; A=144-345.  
 DR Genew; HGNC:6149; ITGAM.  
 DR MIM; 120980; -;  
 DR GO; GO:0008305; C:integrin complex; TAS.  
 DR GO; GO:0007155; P:cell adhesion; TAS.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PRO1185; INTEGRIN.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN ALPHA; 1.  
 DR PROSITE; PS50234; VFPA; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Receptor; Repeat; Signal;  
 KW Transmembrane.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1152 Integrin alpha-M.  
 Query Match 9.5%; Score 145.5; DB 1; Length 1152;  
 Best Local Similarity 26.2%; Pred. No. 0.01; 81; Indels 43; Gaps 12;  
 Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;  
 QY 18 DLVFLIDKSGSVL-HHNNELYYFVQLAKRISPOLRMSFVSTRTGTLTKLTD----- 72  
 Db 150 DIAFLIDGSSIIIPHDRKMEFVST-----VMEQLKSKTLPF-----LMQSEEPRIH 199  
 QY 73 -----RQIRGLEELQKVPDGTVMHBEFERSSEQIYYENRGYRTA-SVITALT 123  
 Db 200 FTFKEPQNNPNPSRLVPIQLL--GRTHATATGIRKVRBELFNTNGARKNAFKILVIT 257  
 QY 124 DGEIHEDLFYYSR--RANRSRDGLATVCGVVDVETGLAR-----IADS--KDHVP 174  
 Db 258 DGEKFGDPLGEDVETPEADRE--GVIRVIYGVGDARSRSEKROELNTIASKPRDHVFQ 314  
 QY 175 VNDGFQALQGIHSHILKSCIEILAAEPSTICAGESFQVAVRGNGFPHA 223  
 Db 315 VNN-FELTKTIQNLREK-----IFALIGTGTGSSSSEHMSQEGFSA 358  
 RESULT 14  
 Q8T5C2 PRELIMINARY; PRT; 453 AA.  
 AC Q8T5C2;  
 DT 01-JUN-2002 (TREMURel. 21, Created)  
 DT 01-JUN-2002 (TREMURel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMURel. 24, Last annotation update)  
 DE Proximal thread matrix protein 1.  
 OS Mytilus galloprovincialis (Mediterranean mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytilloida;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=29156;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;  
 RA Sun C., Lucas J.M., Waite J.H.;  
 RT "collagen-binding matrix proteins from elastomeric extracellular  
 RT byssal fibers." ;  
 RL Biomacromolecules 3:1240-1248(2002).  
 DR EMBL; AY053391; AAL17974.1; -;

```

DR HSP; P20701; IMJN.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VWF_A; 2.
DR Matrix protein.
SQ SEQUENCE 453 AA; 48784 MW; D60497F5C0C51E6D CRC64;

Query Match 9.4%; Score 145; DB 2; Length 453;
Best Local Similarity 25.5%; Pred. No. 0.0036;
Matches 51; Conservative 44; Mismatches 79; Indels 26; Gaps 12;

OY 13 CYGGEFDYFIDKSGSV-----LHMNEIYFVEQLAHKF--ISPO-LRMSFVSTGTT 65
DB 248 CAGADIAFVFDASSINANNPNYQMKRPMKQIVDFRKTGPDGQFANVTADATK 307
OY 66 LMKLTE--DREQIRGLEELQKVLPG--GDTYHGEGERASEQIYENRQG--YRTASV 118
DB 308 QFGKDYSSKADIKGAID---KVSPTIIGQTAIGDLENARLEV-FPNRNGGGRREVQKV 363
OY 119 IIALTDGL--HELDFFSEAREARSDLAIVYCVG-KDFNTQLARTADSDHPEV 175
DB 364 VILLTDQNNNGHKS---PEHSSILRKEGVVAIGVGTFKLSELINIASSEYVF-T 418
OY 176 NDGFQALQGIHSILKSCSI 195
DB 419 TSSFDKSKIMEDVYKLAQM 438

RESULT 15
ITAX HUMAN STANDARD; PRT; 1163 AA.
ID P20702; O81VA6;
AC P20702; O81VA6; 17, Created
DT 01-FEB-1991 (Rel. 14, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Integrin alpha-X precursor (Leukocyte adhesion glycoprotein p150,95
DE alpha chain) (Leukocyte adhesion receptor p150,95) (CD11c) (Lew M5).
CN Name=ITGAX; Synonym=CD11C;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=8616645; PubMed=3327687;
RA Corbi A.L., Miller L.U., O'Connor K., Larson R.S., Springer T.A.;
RT "cDNA cloning and complete primary structure of the alpha subunit of a
RT leukocyte adhesion glycoprotein, p150,95."
RL EMBO J. 6:4023-4028(1987).
[2]
RP SEQUENCE FROM N.A.
RA MEDLINE=90153906; PubMed=2303426;
RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
RT "Genomic structure of an integrin alpha subunit, the leukocyte p150,95
RT molecule."
RL J. Biol. Chem. 265:2782-2788(1990).
[3]
RP ERRATUM.
RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
RL J. Biol. Chem. 265:12750-12751(1990).
[4]
RP SEQUENCE FROM N.A.
RA TISSUE=Blood;
RA MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeebner B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heleh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```

```

RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bosa S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Rahe S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren R.J., Lu X., Gibbs R.A.,
RA Faney J., Helton R., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[5]
RP SEQUENCE OF 20-43.
RX MEDLINE=87167596; PubMed=3549901;
RA Miller L.J., Wiebe M., Springer T.A.;
RT "Purification and alpha subunit N-terminal sequences of human Mac-1
RT and p150,95 leukocyte adhesion proteins."
RL J. Immunol. 138:2381-2383(1987).
CC -1- FUNCTION: Integrin alpha-X/beta-2 is a receptor for fibrinogen. It
CC recognizes the sequence G-P-R in fibrinogen. It mediates cell-cell
CC interaction during inflammatory responses. It is especially
CC important in monocyte adhesion and chemotaxis.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-X
CC associates with beta-2.
CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and
CC granulocytes.
CC -1- DOMAIN: The integrin I-domain (inset) is a VWF domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWF domain.
CC -1- DATABASE: NAME=PROX; NOTE=CD guide CD11c entry;
CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M81695; AA59180.1; -;
DR EMBL; M29165; -; NOT ANNOTATED CDS.
DR EMBL; M29487; AA51620.1; ALT_SEQ.
DR EMBL; M29482; AA51620.1; JOINED.
DR EMBL; M29483; AA51620.1; JOINED.
DR EMBL; M29484; AA51620.1; JOINED.
DR EMBL; M29485; AA51620.1; JOINED.
DR EMBL; M29486; AA51620.1; JOINED.
DR EMBL; BC038237; AAH38237.1; -;
DR PIR; A36584; RWHULC.
DR PDB; 1N3Y; X-ray; A=141-338.
DR Genew; HGNC:6152; ITGAX.
DR MIM; 151510; -;
DR GO; GO:0008305; C:integrin complex; TAS.
DR GO; GO:0004872; F:receptor activity; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR GO; GO:0009887; P:organogenesis; TAS.
DR InterPro; IPR00413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PRO1185; VWFADOMAIN.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.

```

DR PROSITE; PS50234; WMPA; 1.  
 KW 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Polymorphism; Receptor; Repeat;  
 KW Signal; Transmembrane.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1163  
 FT DOMAIN 20 1107  
 FT TRANSMEM 1108 1128  
 FT DOMAIN 1129 1163  
 FT REPEAT 34 87  
 FT REPEAT 88 7  
 FT DOMAIN 165 351  
 FT REPEAT 401 401  
 FT REPEAT 402 453  
 FT REPEAT 455 517  
 FT REPEAT 518 576  
 FT REPEAT 581 633  
 FT CA\_BIND 466 474  
 FT CA\_BIND 530 538  
 FT CA\_BIND 593 601  
 FT SITE 1131 1135  
 FT DISULFID 69 76  
 FT DISULFID 108 126  
 FT DISULFID 655 712  
 FT DISULFID 771 777  
 FT DISULFID 848 863  
 FT DISULFID 998 1022  
 FT DISULFID 1027 1032  
 FT CARBOHYD 61 61  
 FT CARBOHYD 89 89  
 FT CARBOHYD 392 392  
 FT CARBOHYD 697 697  
 FT CARBOHYD 735 735  
 FT CARBOHYD 899 899  
 FT CARBOHYD 939 939  
 FT CARBOHYD 1050 1050  
 FT VARIANT 48 48  
 FT CONFLICT 209 209  
 FT CONFLICT 251 251  
 FT CONFLICT 469 469  
 FT CONFLICT 490 490  
 FT CONFLICT 547 547  
 FT CONFLICT 756 756  
 FT CONFLICT 819 819  
 FT CONFLICT 1161 1163  
 FT STRAND 150 157  
 FT TURN 160 161  
 FT HELIX 164 178  
 FT TURN 179 180  
 FT TURN 183 185  
 FT STRAND 186 193  
 FT STRAND 197 201  
 FT HELIX 203 208  
 FT HELIX 212 216  
 FT TURN 217 218  
 FT STRAND 226 226  
 FT HELIX 228 236  
 FT TURN 237 240  
 FT HELIX 242 244  
 FT TURN 245 245  
 FT TURN 248 249  
 FT STRAND 251 258  
 FT STRAND 263 263  
 FT HELIX 269 278  
 FT TURN 279 280  
 FT STRAND 282 288  
 FT HELIX 290 293  
 FT TURN 296 297  
 FT HELIX 298 304  
 FT HELIX 310 312  
 FT STRAND 313 316  
 FT HELIX 319 325

Integrin alpha-X.  
 Extracellular (Potential).  
 Potential.  
 Cytoplasmic (Potential).  
 FG-GAP 1.  
 FG-GAP 2.  
 WMPA.  
 FG-GAP 3.  
 FG-GAP 4.  
 FG-GAP 5.  
 FG-GAP 6.  
 FG-GAP 7.  
 Potential.  
 Potential.  
 Potential.  
 GFPR motif.  
 By similarity.  
 By similarity.  
 By similarity.  
 By similarity.  
 By similarity.  
 By similarity.  
 By similarity.  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 N-linked (GlcNAc. . .) (Potential).  
 W -> R (in dbSNP:11574633).  
 /FTid=VAR\_018672.  
 T -> S (in Ref. 4).  
 T -> A (in Ref. 4).  
 T -> S (in Ref. 4).  
 G -> A (in Ref. 2).  
 E -> K (in Ref. 4).  
 D -> L (in Ref. 1).  
 I -> V (in Ref. 4).  
 SEK -> TPVYQDNV (in Ref. 4).

FT HELIX 326 334  
 FT TURN 335 335  
 SQ SEQUENCE 1163 AA; 127886 MW; 83658A13B5C5D8F CRC64;  
 Query Match 9.3%; Score 143.5; DB 1; Length 1163;  
 Best Local Similarity 24.8%; Pred. No. 0.015;  
 Matches 55; Conservative 43; Mismatches 89; Indels 35; Gaps 10;  
 QY 18 DLYFILDKSGSV-LHHNNEIYFVEQLAHKEIPQLMSPFVSTRTGLTKLTEDRE-- 74  
 DB 151 DIVFLIDSGSGSISRNPAIMNFVRAVISQQRSPQPSLMQFENKQHTHTFEFRRTS 210  
 QY 75 ---QIRQGLBELQVLPQSDTYMHEGPRASQIYYENRQGYTAS-VIALLTGEELHED 130  
 DB 211 NPLSLASVHQLQ-----GFTYTATAIQNVVHRLFHASYGARRDATKILIVTDGKKEGD 265  
 QY 131 LFPYSEHANSRRLGAIYVCVGV-----KDFNETQLANIAD--SKDHVPVNDG 178  
 DB 266 SLDYKD-VIPMDAAGIIRYAIQVGLAFQNRNSKELND-----TASKPSOEHTFVED- 318  
 QY 179 FOALQGIHSHILKKSCEIILAEPSTICAGSEFQVVRNGVF 220  
 DB 319 FDLAKDQIONQKXK-----FAIEGTETSSSSFFELMAQEGF 356

Search completed: June 13, 2005, 20:03:13  
 Job time : 102.03 secs

**THIS PAGE BLANK (USPTO)**



us-09-970-076-1\_copy\_104\_1207.rge

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd

OM nucleic - nucleic search, using sw model

Run on: June 14, 2005, 11:31:32 ; Search time 5146 Seconds  
(without alignments)  
10395.367 Million cell updates/sec

Title:	US-09-970-076-1_COPY_104_1207
Perfect score:	1104
Sequence:	1 atgccacgcgscgcggag.....gtgaggaaataataaaaa 1104

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl: \*

```

1:  gb_da:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pac:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_stb:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1104	100.0	1414	6	AX458366	AX458366 Sequence
2	1104	100.0	1414	6	AF441380	AF441380 Homo sapi
3	1090.8	98.8	5540	6	AX393246	AX393246 Sequence
4	1090.8	98.8	5540	6	AX393301	AX393301 Sequence
5	1090.8	98.8	5540	6	AX458370	AX458370 Sequence
6	1090.8	98.8	5540	9	AF279145	AF279145 Sequence
7	949.4	86.0	2112	6	AX458372	AX458372 Homo sapi
8	949.4	86.0	2112	6	AX458372	AX458372 Sequence
9	853.4	77.3	5220	6	BC012074	BC012074 Homo sapi
10	853.4	77.3	5220	6	AX393265	AX393265 Sequence
11	853.4	77.3	5220	6	AX393370	AX393370 Sequence
12	803.8	72.8	1609	6	AR338517	AR338517 Mus muscu
13	802.2	72.7	1718	6	CQ498433	AR338517 Sequence
14	787	71.3	1436	6	BD156323	CQ498433 Sequence
15	787	71.3	1436	6	AX458374	BD156323 Primer fo
16	787	71.3	1436	6	AX458374	AX458374 Sequence
17	787	71.3	1436	6	AX676800	AX676800 Sequence
18	657.8	60.5	4739	10	AK001463	AK001463 Homo sapi
19	542.8	49.2	4007	9	BC057043	AK001463 Mus muscu
					AK055429	BC057043 Homo sapi
						AK055429 Homo sapi

20	349	11.6	356	6	CQ426476	Sequence
21	348	11.5	355	6	CQ415556	Sequence
22	346.2	11.4	2234	6	AX747144	Sequence
23	346.2	11.4	2234	9	AK0911721	Homo sapi
24	346	11.3	346	6	CQ433315	Sequence
25	344.6	11.2	1464	6	BD272493	Secreted
26	344.6	11.2	1467	6	BD272541	Secreted
27	344.6	11.2	1470	9	AY233452	Homo sapi
28	344.6	11.2	3677	6	BD272492	Secreted
29	344.6	11.2	4348	6	CQ414929	Sequence
30	344.6	11.2	4350	6	BD272540	Secreted
31	343	11.1	3677	6	BD272508	Secreted
32	343	11.1	3677	6	BD272509	Secreted
33	343	11.1	3677	6	BD272510	Secreted
34	342.8	11.1	1461	6	BD272512	Secreted
35	341.8	11.0	3501	6	BD272513	Secreted
36	340.2	10.8	1458	6	BD272553	Secreted
37	340.2	10.8	1461	6	BD272503	Secreted
38	340.2	10.8	3501	6	BD272502	Secreted
39	340.2	10.8	3501	6	BD272511	Secreted
40	340.2	10.8	3776	10	BC076595	Mus muscu
41	338.6	10.7	3501	6	BD272512	Secreted
42	307.4	27.8	1492	6	AR542049	Sequence
43	277.6	25.1	464	6	BD148180	Primer fo
44	277.6	25.1	464	6	AX868118	Sequence
45	274	24.8	553	6	CQ715841	Sequence

## ALIGNMENTS

RESULT	1
AX458366	
LOCUS	AX458366 1414 bp DNA
DEFINITION	Sequence 1 from Parent W00246228.
ACCESSION	AX458366
VERSION	AX458366.1 GI:21725038
KEYWORDS	.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
PAT	08-JUL-2007

REFERENCE  
1  
Young, J.A., Bradley, K.A., Collier, R.J. and Mogridge, J.S  
AUTHORS  
TITLE  
Receptor for b. Anthracis toxin  
JOURNAL  
Patent: WO 0246228-A 1 13-JUN-2002;

WISCONSIN ALUMNI RESEARCH FOUNDATION (US	location/Qualifiers
FEATURES	

**Source**

**CDS**

```

/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
104..1210
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD38349.1"
/db_xref="GI:21725039"
/translation="MATEIRRALGIGHOWLSLATLVJTAGGGGEREDGSPACVGGF
LYFLIDSGSVLAHMIAYFVQQLAHKFLSPDLRSLFVFSFRTGLTMKLTDEKQ
RQGLELQQLVPGCDYTRGDFRASEQQLYENRQCFRASYVIALTQGLIEDHFLPK
SERENRNRDGLADYICVGYKDNFQRLAIRSKQHPVNDGQALQGLIHSLIKF
SYCLIALAEPTQAGSSPVVIRVGRFHARVNVVIVCSFPIINDSVLTINKEPVEVL
TYLCEAPRIKEVGNKALQVSNRQDGLSFSSVYITTHRCDSGLIALALLIPLLL
ALLALLMFWPLCTVILIIKEVNPAAESENKTK"

```

## ORIGIN

Query Match	100.0%;	Score 1104;	DB 6;	Length 1414;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1104; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	Db
1	1
104	104
163	163

```

QY 61 CTGCTGCTCATCTGCGCGCGGCGAAGGAGATGAGGATGCGGCTGCTAC 120
DB 164 CTGGGTCTCATCTGCGCGCGGCGAAGGAGATGAGGATGCGGCTGCTAC 223
QY 121 GCGCGATTTGACCTGACTTCTATTTTGGACAAATCAGAGATGCTGACCACTGGAAAT 180
DB 224 GCGCGATTTGACCTGACTTCTATTTTGGACAAATCAGAGATGCTGACCACTGGAAAT 283
QY 101 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCACAGTTGAGATG 240
DB 284 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCACAGTTGAGATG 343
QY 241 TCCCTTATTTGTTTTCTCCACCCGAGAAACAACTTATGAACTGACAGAAACAGAGAA 300
DB 344 TCCCTTATTTGTTTTCTCCACCCGAGAAACAACTTATGAACTGACAGAAACAGAGAA 403
QY 301 CAAATCCGTCAAGGCGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360
DB 404 CAAATCCGTCAAGGCGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 463
QY 361 CATGAAGATTTGAAGGCGCGAGAGAGATTTATTTGAAAAAGACAAAGGCTACAG 420
DB 464 CATGAAGATTTGAAGGCGCGAGAGAGATTTATTTGAAAAAGACAAAGGCTACAG 523
QY 421 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTTCATGAAATCTCTTTTCTAT 480
DB 524 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTTCATGAAATCTCTTTTCTAT 583
QY 481 TCAGAGAGAGAGGCTTATATGCTTCGAGATCTTGGTCAATTTGTTTACTGTGTGTG 540
DB 584 TCAGAGAGAGAGGCTTATATGCTTCGAGATCTTGGTCAATTTGTTTACTGTGTGTG 643
QY 541 AAAAGATTTCAATGAGACACAGCTGGCGCGAGTTGGAGACAGTATGATCATGTGTTTCC 600
DB 644 AAAAGATTTCAATGAGACACAGCTGGCGCGAGTTGGAGACAGTATGATCATGTGTTTCC 703
QY 601 GTGAATGACGCGCTTCAAGCTCTGCAAGGAGCATCATCACTCAATTTTGAAGAGTCTGCG 660
DB 704 GTGAATGACGCGCTTCAAGCTCTGCAAGGAGCATCATCACTCAATTTTGAAGAGTCTGCG 763
QY 661 ATCGAAATTTCTGACGCTGAGAACATCCACATATGTGACAGAGATCATTTCAATTTGTC 720
DB 764 ATCGAAATTTCTGACGCTGAGAACATCCACATATGTGACAGAGATCATTTCAATTTGTC 823
QY 721 GTGAGAGGAAAGCGCTTCCGACATGCGCGCAAGTGAAGAGGCTCTGACGTTCAAG 780
DB 824 GTGAGAGGAAAGCGCTTCCGACATGCGCGCAAGTGAAGAGGCTCTGACGTTCAAG 883
QY 781 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTTCTGTGAAAGACACTTATTTACTG 840
DB 884 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTTCTGTGAAAGACACTTATTTACTG 943
QY 841 TGTCCAGCGCTTATCTTAAAGAGTGGCATGAAAGTGCATCTCAGGTGACATGAAAC 900
DB 944 TGTCCAGCGCTTATCTTAAAGAGTGGCATGAAAGTGCATCTCAGGTGACATGAAAC 1003
QY 901 GATGCGCTCTCTTTATCTCAGTCTGTCATCATCAACACACACTGTTCTGACGAT 960
DB 1004 GATGCGCTCTCTTTATCTCAGTCTGTCATCATCAACACACACTGTTCTGACGAT 1063
QY 961 TCCATCTTGCGCATGCGCTGCTGATCTGTTCTGTGCTTGAAGCCCTGCGCTCTCTGCG 1020
DB 1064 TCCATCTTGCGCATGCGCTGCTGATCTGTTCTGTGCTTGAAGCCCTGCGCTCTCTGCG 1123
QY 1021 TGGTCTTGCGCCCTCTGCTGCACTGTGATATCAAGAGAGTCCCTTCAACCCCTGCGAG 1080
DB 1124 TGGTCTTGCGCCCTCTGCTGCACTGTGATATCAAGAGAGTCCCTTCAACCCCTGCGAG 1183
QY 1081 GAGAGTGAGAAATTAATTAATA 1104
DB 1184 GAGAGTGAGAAATTAATTAATA 1207

```

```

RESULT 2
LOCUS AF421380 1414 bp mRNA linear PRI 13-NOV-2001
DEFINITION Homo sapiens anthrax toxin receptor mRNA, complete cds.
ACCESSION AF421380
VERSION AF421380.1 GI:16566412
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Bradley,K.A., Mougidge,J., Mougidge,M., Collier,R.J. and Young,J.A.
TITLE Identification of the cellular receptor for anthrax toxin
JOURNALS Nature 414 (6860), 225-229 (2001)
MEDLINE 21557240
PUBMED 11700562
REMARK http://www.nature.com
2 (bases 1 to 1414)
REFERENCE Bradley,K.A., Mougidge,J., Mougidge,M., Collier,R.J. and Young,J.A.
AUTHORS Direct Submission
TITLE Submitted (19-SRP-2001) Department of Oncology, University of
JOURNALS Wisconsin-Madison, 1400 University Ave., Madison, WI 53706, USA
FEATURES
source
1..1414
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="Hela"
1..103
104..1210
/codon_start=1
/feature="ATG"
/protein_id="P4126486.1"
/protein="anthrax toxin receptor"
/db_xref="GI:16566413"
/translation="MATERRALGIGFOWLSLATVLICAGGGRREDGPGACGFD
LPLTLDKSGSVLHHMNIYFVQLAKFPLSPOLRMSPIVSTRTGLMLKLTEDREI
ROGLELEOKYLPQGDITMHEGFERASEQIYVENOGRTASVIALDGEHLDFPY
SEREANSRLGAILVCVGVKDFNETDLAIADSKHVPVNDGFOALOGIHSILK
SCIEIILAPBSTICAGESPOVYVNGNFRHARVDRVLCSPKINDSVLANKPSVD
TYLICPAPIIKERGMKRALDVSMDGSIPLSSSVITTTTCHSDSIIAAILLIFLL
ALALWFWPLCCVTIIKEVPPPEESSENKIK"
1211..1414
ORIGIN
3'UTR
Query Match 100.0%; Score 1104; DB 9; Length 1414;
Best Local Similarity 100.0%; Pied. No. 0;
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCCACGCGCGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60
DB 104 ATGCCACGCGCGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 163
QY 61 CTGCTGCTCATCTGCGCGCGGCGAAGGAGATGAGGATGCGGCTGCTAC 120
DB 164 CTGCTGCTCATCTGCGCGCGGCGAAGGAGATGAGGATGCGGCTGCTAC 223
QY 121 GCGCGATTTGACCTGACTTCTATTTTGGACAAATCAGAGATGCTGACCACTGGAAAT 180
DB 224 GCGCGATTTGACCTGACTTCTATTTTGGACAAATCAGAGATGCTGACCACTGGAAAT 283
QY 241 TCCCTTATTTGTTTTCTCCACCCGAGAAACAACTTATGAACTGACAGAAACAGAGAA 300
DB 344 TCCCTTATTTGTTTTCTCCACCCGAGAAACAACTTATGAACTGACAGAAACAGAGAA 403
QY 301 CAAATCCGTCAAGGCGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360

```

Db 404 CAATCCGTCAGAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 463  
 Qy 361 CATGAAGATTGAAAAGGCGCAATGAGACAGATTATATGAAAACAGACAGGTAACAG 420  
 Db 464 CATGAAGATTGAAAAGGCGCAATGAGACAGATTATATGAAAACAGACAGGTAACAG 523  
 Qy 421 AAGCCAGGTCATCTTCTGCTTGTGATGAGAACTCCATGAAAGATCTCTTTTCTAT 480  
 Db 524 ACAGCCAGGTCATCTTCTGCTTGTGATGAGAACTCCATGAAAGATCTCTTTTCTAT 583  
 Qy 481 TCAGAGAGGAGGCTTATAGGTCTGAGATCTTGTGCAATTGTTACTGTGTGTGTG 540  
 Db 584 TCAGAGAGGAGGCTTATAGGTCTGAGATCTTGTGCAATTGTTACTGTGTGTGTG 643  
 Qy 541 AAAAGATTCAATGAGACACAGCTGGCCCGAATGCGAGACATGAGATCATGTGTTCCC 600  
 Db 644 AAAAGATTCAATGAGACACAGCTGGCCCGAATGCGAGACATGAGATCATGTGTTCCC 703  
 Qy 601 GTGAATGAGGCTTTCAGGCTCTGCAAGGACATCTCACTCAATTTTGAAGATCTGCG 660  
 Db 704 GTGAATGAGGCTTTCAGGCTCTGCAAGGACATCTCACTCAATTTTGAAGATCTGCG 763  
 Qy 661 ATCGAATTTCTAGCAGCTGACATCCATCATATGTGAGAGAGTCAATTTGAAGTTCG 720  
 Db 764 ATCGAATTTCTAGCAGCTGACATCCATCATATGTGAGAGAGTCAATTTGAAGTTCG 823  
 Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCCCAAGTGAACAGGCTCTCTGACAGTTCAAG 780  
 Db 824 GTGAGAGAAACGGCTTCCGACATGCGCCCAAGTGAACAGGCTCTCTGACAGTTCAAG 883  
 Qy 781 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTCTGTGAAAGACATTAATTAAG 840  
 Db 884 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTCTGTGAAAGACATTAATTAAG 943  
 Qy 841 TGTCCAGGCTCTATCTTAAAGAGTGGCATGAAAGCTGCACTCAGGTGACATGAAAC 900  
 Db 944 TGTCCAGGCTCTATCTTAAAGAGTGGCATGAAAGCTGCACTCAGGTGACATGAAAC 1003  
 Qy 901 GATGCGCTCTCTTTATCTCCAGTTCTGTGATCATCAACACACACACACTGTTCTGACGGT 960  
 Db 1004 GATGCGCTCTCTTTATCTCCAGTTCTGTGATCATCAACACACACACACTGTTCTGACGGT 1063  
 Qy 961 TCCATCTGCGCATGCGCTGCTGATCTGTCTGCTGCTGAGCCCTGCTCTCTGCG 1020  
 Db 1064 TCCATCTGCGCATGCGCTGCTGATCTGTCTGCTGCTGAGCCCTGCTCTCTGCG 1123  
 Qy 1021 TGGTCTGCGCCCTCTGCTGCACTGATATCAAGAGAGTCCCTCCACCCCTGCGAG 1080  
 Db 1124 TGGTCTGCGCCCTCTGCTGCACTGATATCAAGAGAGTCCCTCCACCCCTGCGAG 1183  
 Qy 1081 GAGAGTGAGAAATTAATAAATAA 1104  
 Db 1184 GAGAGTGAGAAATTAATAAATAA 1207

RESULT 3  
 LOCUS AX393246 5540 bp DNA linear PAT 23-MAR-2002  
 DEFINITION Sequence 176 from Patent W00210217.  
 ACCESSION AX393246  
 VERSION AX393246.1 GI:19701296  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE  
 1 St Croix, B., Kinzler, K.W. and Vogelstein, B.  
 TITLE Endothelial cell expression patterns  
 JOURNAL Patent: WO 0210217-A 176 07-FEB-2002;  
 The Johns Hopkins University (US)  
 FEATURES  
 Location/Qualifiers  
 1..5540

/organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"

ORIGIN  
 Query Match 98.8%; Score 1090.8; DB 6; Length 5540;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGCCACGGCGGAGAGAGAGCCCTCGGCAATCGGCTTCCAGTGGCTCTCTTTGGCACT 60  
 Db 144 ATGCCACGGCGGAGAGAGAGCCCTCGGCAATCGGCTTCCAGTGGCTCTCTTTGGCACT 203  
 Qy 61 CTGGTCTCATCTGCGCGCGGCAAGGAGAGAGAGATGGGGGTCCAGCCCTGCTAC 120  
 Db 204 CTGGTCTCATCTGCGCGCGGCAAGGAGAGAGAGATGGGGGTCCAGCCCTGCTAC 263  
 Qy 121 GCGGATTTGACCTGTACTTCAATTTTGAACAATAGAGAGTGTGCTGACCACTGGAAT 180  
 Db 264 GCGGATTTGACCTGTACTTCAATTTTGAACAATAGAGAGTGTGCTGACCACTGGAAT 323  
 Qy 181 GAATCTATTAATTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 240  
 Db 324 GAATCTATTAATTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 383  
 Qy 241 TCTTTATTTGTTTCTCCACCGAGAAACAATTAAGAACTGACAGAAAGACAGAGAA 300  
 Db 384 TCTTTATTTGTTTCTCCACCGAGAAACAATTAAGAACTGACAGAAAGACAGAGAA 443  
 Qy 301 CAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 360  
 Db 444 CAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 503  
 Qy 361 CATGAAGATTGAAAGGCGCAGTGAAGATTAATATGAAAACAGCAAGGGTACAGG 420  
 Db 504 CATGAAGATTGAAAGGCGCAGTGAAGATTAATATGAAAACAGCAAGGGTACAGG 563  
 Qy 421 ACAGCCAGCTCATCTTGTGATCTGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
 Db 564 ACAGCCAGCTCATCTTGTGATCTGATGAGAACTCCATGAAGATCTCTTTTCTAT 623  
 Qy 481 TCAGAGAGAGGCTTAATAGGTCTGAGATCTTGGTGAATTTTACTGTGTGTGTG 540  
 Db 624 TCAGAGAGAGGCTTAATAGGTCTGAGATCTTGGTGAATTTTACTGTGTGTGTG 683  
 Qy 541 AAAAGTTCAATGAGACACAGCTGGCCCGAATGCGGACATTAAGAGATCATGTGTTCC 600  
 Db 684 AAAAGTTCAATGAGACACAGCTGGCCCGAATGCGGACATTAAGAGATCATGTGTTCC 743  
 Qy 601 GTGAATGAGGCTTTCAGGCTCTGCAAGGACATCACTCAATTTTGAAGAGTCTGCG 660  
 Db 744 GTGAATGAGGCTTTCAGGCTCTGCAAGGACATCACTCAATTTTGAAGAGTCTGCG 803  
 Qy 661 ATCGAATTTAGAGAGTGAACATCACTCAATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
 Db 804 ATCGAATTTAGAGAGTGAACATCACTCAATATGTGAGAGAGTCAATTTCAAGTTGTC 863  
 Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGAAAGTGAACAGGTCCTTGTGAGCTTCAAG 780  
 Db 864 GTGAGAGAAACGGCTTCCGACATGCGCGAAAGTGAACAGGTCCTTGTGAGCTTCAAG 923  
 Qy 781 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTCTGTGAGAGATCTTAATTAAG 840  
 Db 924 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTCTGTGAGAGATCTTAATTAAG 983  
 Qy 841 TGTCCAGGCTCTATCTTAAAGAGTGGCATGAAAGCTGCACTCCAGTCCAGCATGAAC 900  
 Db 984 TGTCCAGGCTCTATCTTAAAGAGTGGCATGAAAGCTGCACTCCAGTCCAGCATGAAC 1043  
 Qy 901 GATGCGCTCTCTTTATCTCCAGTTCTGTGATCATCAACACACACACTGTTCTGACGGT 960  
 Db 1044 GATGCGCTCTCTTTATCTCCAGTTCTGTGATCATCAACACACACACTGTTCTGACGGT 1103

Qy 961 TCACATCTGGGCGCATGCGCCCTGCTGATTCCTGTTCTGCTCTAGCCCTGCTCTCTGG 1020  
Db 1104 TCACATCTGGGCGCATGCGCCCTGCTGATTCCTGTTCTGCTCTAGCCCTGCTCTCTGG 1163  
Qy 1021 TGGTCTTGGCCCTCTGCTGCACTGATTAATCAAGAGAGGTCCCTCCACCCCTGCGAG 1080  
Db 1164 TGGTCTTGGCCCTCTGCTGCACTGATTAATCAAGAGAGGTCCCTCCACCCCTGCGAG 1223  
Qy 1081 GAGAGTGAGAAA 1094  
Db 1224 GAGAGTGAGAAA 1237

RESULT 4  
AX393301 5540 bp DNA linear PAT 23-MAR-2002  
LOCUS Sequence 231 from Patent WO0210217.  
DEFINITION AX393301  
ACCESSION AX393301 GI:19701322  
VERSION AX393301.1 GI:19701322  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 St Croix, B., Kinzler, K.W. and Vogelstein, B.  
Endothelial cell expression patterns  
Patent: WO 0210217-A 231 07-FEB-2002;  
The Johns Hopkins University (US)  
FEATURES  
Source  
1..5540  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 98.8%; Score 1090.8; DB 6; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ATGGCCAGCGCGGAGCGGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTTTGGCACT 60  
Db 144 ATGGCCAGCGCGGAGCGGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTTTGGCACT 203  
Qy 61 CTGGTCTCATCTGGCGCCGCGGCAAGGGGACGCGAGAGATGGGGGTCCACCTGCTAC 120  
Db 204 CTGGTCTCATCTGGCGCGCGGCAAGGGGACGCGAGAGATGGGGGTCCACCTGCTAC 263  
Qy 121 GCGCGATTGACCTGTACTTTTGGCAAAATCAGAAAGTGTGTGCACTGCACTGAAT 180  
Db 264 GCGCGATTGACCTGTACTTTTGGCAAAATCAGAAAGTGTGTGCACTGCACTGAAT 323  
Qy 181 GAAATCTATTACTTTGTGAGACAGTTGGCTCAAAATTCATGAGCCCAAGTTGGAATG 240  
Db 324 GAAATCTATTACTTTGTGAGACAGTTGGCTCAAAATTCATGAGCCCAAGTTGGAATG 383  
Qy 241 TCTTTATTGTTTTCACCCGAGGAAACAATTATGAAAATGACAGAGACAGAGA 300  
Db 384 TCTTTATTGTTTTCACCCGAGGAAACAATTATGAAAATGACAGAGACAGAGA 443  
Qy 301 CAAATCCGTCAGAGGCTTGAAGAACTCCGAAAGTTCTGCGAGAGAGACCTTAATG 360  
Db 444 CAAATCCGTCAGAGGCTTGAAGAACTCCGAAAGTTCTGCGAGAGAGACCTTAATG 503  
Qy 361 CATTGAAGATTGAAAAGGCGCATGAGAGATTATTAAGAAAACGACAAAGGTTACAG 420  
Db 504 CATTGAAGATTGAAAAGGCGCATGAGAGATTATTAAGAAAACGACAAAGGTTACAG 563  
Qy 421 ACAGCCAGCGCTATCTGCTTTGACTGATGAGAACTCCATGAAGATCTTTTTCAT 480  
Db 564 ACAGCCAGCGCTATCTGCTTTGACTGATGAGAACTCCATGAAGATCTTTTTCAT 623  
Qy 481 TCAGAGAGGAGGCTTAATAGGTCTCGAGATCTTGGTCAATGTTTACTGTGTGTGTG 540

Db 624 TCAGAGAGGAGGCTTAATAGGTCTCGAGATCTTGGTCAATGTTTACTGTGTGTGTG 683  
Qy 541 AAAGATTTCATAGACACAGACTGGCCGGATTGCGGACAGTAAGATCATGTTTCCC 600  
Db 684 AAAGATTTCATAGACACAGACTGGCCGGATTGCGGACAGTAAGATCATGTTTCCC 743  
Qy 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAGTCTGC 660  
Db 744 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAGTCTGC 803  
Qy 661 ATGAAATTCAGAGCGTGAACCATCCACATATGTCAGAGAGTCAATTTCAAGTTGTC 720  
Db 804 ATGAAATTCAGAGCGTGAACCATCCACATATGTCAGAGAGTCAATTTCAAGTTGTC 863  
Qy 721 GTGAGAGAAAAGCGCTTCCGACATGCCCGCAACGTGACAGAGGTCTCTGAGCTTCAAG 780  
Db 864 GTGAGAGAAAAGCGCTTCCGACATGCCCGCAACGTGACAGAGGTCTCTGAGCTTCAAG 923  
Qy 781 ATCAATGACTGCTGACATCTCAATGAGAGGCTTTTCTGTGAAGACATTAATTTACTG 840  
Db 924 ATCAATGACTGCTGACATCTCAATGAGAGGCTTTTCTGTGAAGATTAATTTACTG 983  
Qy 841 TGTCCAGCGCTATCTTAAGAAAGTTGGCATGAAGCTGACCTCCAGTCCAGATGAC 900  
Db 984 TGTCCAGCGCTATCTTAAGAAAGTTGGCATGAAGCTGACCTCCAGTCCAGATGAC 1043  
Qy 901 GATGGCTCTCTTTTATCTCCAGTTCTGTGATCATCAACACACACACTGTTTCAAGGT 960  
Db 1044 GATGGCTCTCTTTTATCTCCAGTTCTGTGATCATCAACACACACTGTTTCAAGGT 1103  
Qy 961 TCACATCTGGGCGCATGCGCCCTGCTGATTCCTGTTCTGCTCTAGCCCTGCTCTCTGG 1020  
Db 1104 TCACATCTGGGCGCATGCGCCCTGCTGATTCCTGTTCTGCTCTAGCCCTGCTCTCTGG 1163  
Qy 1021 TGGTCTTGGCCCTCTGCTGCACTGATTAATCAAGAGAGGTCCCTCCACCCCTGCGAG 1080  
Db 1164 TGGTCTTGGCCCTCTGCTGCACTGATTAATCAAGAGAGGTCCCTCCACCCCTGCGAG 1223  
Qy 1081 GAGAGTGAGAAA 1094  
Db 1224 GAGAGTGAGAAA 1237

RESULT 5  
AX458370 5540 bp DNA linear PAT 08-JUL-2002  
LOCUS Sequence 5 from Patent WO0246228.  
DEFINITION AX458370  
ACCESSION AX458370  
VERSION AX458370.1 GI:21725040  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 Young, J.A., Bradley, K.A., Collier, R.J. and Mowridge, J.S.  
Receptor for b. Anthracis toxin  
Patent: WO 0246228-A 5 13-JUN-2002;  
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)  
FEATURES  
Source  
1..5540  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
144..1838  
/note="unamed protein product"  
/codon\_start=1  
/protein\_id="CAD38350.1"  
/db\_xref="GI:21725041"  
/translation="MATVERALIGFOMSLATLVILICAGGGRREDGPACYGFD  
LYFLIDSGSVLHHMNIYVFEOLAKRFLSPOLRMSPIVSTGTTLMKLDEDEOI  
RQGLEIOLKVLPGSDTYMHBEFERASEQIYIENQGRITASVIALITDGEIHEDLFFY

SEBANSRDLGAIYCVGKDFNETQLARIADSKDHYFVNDGFOALQGIHLSILKK  
SCIEILAEPTICAGBSFQVVRNGFPHANVRVLCSPKINSVLINPEFVED  
TYLILPAILKEIKGKAAALQVMDNLSITSSVITTTGSDGSLIALILIFLL  
ALALIMFMPLECCVTIKEYPPPAESEEDDDGPKKMTVTNVAASYGAGVGITK  
RMEVWGEKSTEBGALKAKARVMPQOEYFEPERNLNMMNRSPSPKXYSPI  
KGDALMVLIRKGYDRSVNRPQEDTRCINFRVKNQPAKPLNNAVHTSSPP  
APVTPPPAPHCPPPPSAFPPIPSPESTLPPPPQAPPPNAPPPSPRPPPSV"

## ORIGIN

Query Match 98.8%; Score 1090.8; DB 6; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCAGCGCGGAGGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTTTGGCCACT 60  
DB 144 ATGGCCAGCGCGGAGGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTTTGGCCACT 203  
QY 61 CTGGTGCTCATCTGGCCCGGCGCAAGGGGAGCGAGGAGATGGGGGTCCAGCTGTAC 120  
DB 204 CTGGTGCTCATCTGGCCCGGCGCAAGGGGAGCGAGGAGATGGGGGTCCAGCTGTAC 263  
QY 121 GGGCGATTGACCTGACTTCAATTTTGA CAATCAGAAAGTGTGTGACCACTGGAAT 180  
DB 264 GGGCGATTGACCTGACTTCAATTTTGA CAATCAGAAAGTGTGTGACCACTGGAAT 323  
QY 181 GAAATCTAATCTTGTGGAGCAAGTGGCTCAAAATTCATCAGCCCAAGTTGAGAATG 240  
DB 324 GAAATCTAATCTTGTGGAGCAAGTGGCTCAAAATTCATCAGCCCAAGTTGAGAATG 383  
QY 241 TCCTTTATTTGTTTCTCCACCCGAGAGAACACTTATGAAA CTGACAGAGACAGAGA 300  
DB 384 TCCTTTATTTGTTTCTCCACCCGAGAGAACACTTATGAAA CTGACAGAGACAGAGA 443  
QY 301 CAATCCGTCAGAGGCTGAGAGACTCCAGAAAGTTGCGCGAGGAGAGCACTTAATG 360  
DB 444 CAATCCGTCAGAGGCTGAGAGACTCCAGAAAGTTGCGCGAGGAGAGCACTTAATG 503  
QY 361 CATGAGAGATTTGAAAAGGCCAGTAGAGCATTTTATGAAAAGACAGAGGTACAGG 420  
DB 504 CATGAGAGATTTGAAAAGGCCAGTAGAGCATTTTATGAAAAGACAGAGGTACAGG 563  
QY 421 ACAGCCAGGTCATCATTTGCTTGA CTGATGAGAGAACTCCATGAAGATCTTTTTCAT 480  
DB 564 ACAGCCAGGTCATCATTTGCTTGA CTGATGAGAGAACTCCATGAAGATCTTTTTCAT 623  
QY 481 TCAGAGAGGAGGCTTAATAGCTCTGAGATCTTGGTGCAATTTGTTTCTGTTGGTGTG 540  
DB 624 TCAGAGAGGAGGCTTAATAGCTCTGAGATCTTGGTGCAATTTGTTTCTGTTGGTGTG 683  
QY 541 AAAAGATTTCAATGAGACACAGCTGGCCGAGATTGGGACAGTAPAGATCATGTGTTCC 600  
DB 684 AAAAGATTTCAATGAGACACAGCTGGCCGAGATTGGGACAGTAPAGATCATGTGTTCC 743  
QY 601 GTGAATGACGAGCTTCAGGCTCTGCAAGGCATCATCACTCAATTTTGAAGAGTCTGTC 660  
DB 744 GTGAATGACGAGCTTCAGGCTCTGCAAGGCATCATCACTCAATTTTGAAGAGTCTGTC 803  
QY 661 ATCGAAATTTTGAAGAGTGAACCAATCCATATGTGACAGAGATCATTTCAATGTGTC 720  
DB 804 ATCGAAATTTTGAAGAGTGAACCAATATGTGACAGAGATCATTTCAATGTGTC 863  
QY 721 GTGAGAGAGAAAGCGCTTCGACATCCGCGCAAGTGAAGAGGCTCTGACAGTTCAGG 780  
DB 864 GTGAGAGAGAAAGCGCTTCGACATCCGCGCAAGTGAAGAGGCTCTGACAGTTCAGG 923  
QY 781 ATCAATGACCTGGGTCACTCAATGAGAGCCCTTTTCTGTGGAAGACATTAATTACTG 840  
DB 924 ATCAATGACCTGGGTCACTCAATGAGAGCCCTTTTCTGTGGAAGATTAATTACTG 983  
QY 841 TGTCCAGGCGCTATTTTAAAGAAAGTTGGCATGAAAGCTGCATCTCAGCTCAGCATGAC 900  
DB 984 TGTCCAGGCGCTATTTTAAAGAAAGTTGGCATGAAAGCTGCATCTCAGCTCAGCATGAC 1043

QY 901 GATGCGCTCTCTTTATCTCAGTTCATGATCATCACACACACACTGTTCTGACGAT 960  
DB 1044 GATGCGCTCTCTTTATCTCAGTTCATGATCATCACACACACACTGTTCTGACGAT 1103  
QY 961 TCCATCTCGGCATGAGCCCTGTGTATCTCTGTTCTGTTCTTACAGCCCTGCTCTG 1020  
DB 1104 TCCATCTCGGCATGAGCCCTGTGTATCTCTGTTCTGTTCTTACAGCCCTGCTCTG 1163  
QY 1021 TGGTCTGCGCCCTCTGCTGACCTGTGATTTCAAGAGAGTCCCTCCACCCCTGCGAG 1080  
DB 1164 TGGTCTGCGCCCTCTGCTGACCTGTGATTTCAAGAGAGTCCCTCCACCCCTGCGAG 1223  
QY 1081 GAGACTGAGAAA 1094  
DB 1224 GAGACTGAGAAA 1237  
RESULT 6  
AF279145  
LOCUS  
DEFINITION Homo sapiens tumor endothelial marker 8 precursor (TEM8) mRNA,  
AF279145  
ACCESSION AF279145  
VERSION AF279145.2 GI:14017380  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 5540)  
St. Croix,B., Rego,C., Velculescu,V., Traverso,G., Romans,K.B.,  
Montgomery,E., Lal,A., Riggin,G.J., Lengauer,C., Vogelstein,B. and  
Kinzler,K.W.  
Genes expressed in human tumor endothelium  
Science 289 (5482), 1197-1202 (2000)  
PUBMED  
20407466  
10947988  
REFERENCE  
2 (bases 1 to 5540)  
St. Croix,B., Vogelstein,B. and Kinzler,K.W.  
Direct Submission  
Submitted (16-JUN-2000) Johns Hopkins Oncology Center, Johns  
Hopkins University, 1650 Orleans Street, Baltimore, MD 21231, USA  
3 (bases 1 to 5540)  
St. Croix,B., Vogelstein,B. and Kinzler,K.W.  
Direct Submission  
Submitted (09-MAY-2001) Johns Hopkins Oncology Center, Johns  
Hopkins University, 1650 Orleans Street, Baltimore, MD 21231, USA  
REMARK  
COMMENT  
On May 9, 2001 this sequence version replaced gi:19857405.  
FEATURES  
source  
1..5540  
location/Qualifiers  
1..5540  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
1..5540  
/gene="TEM8"  
144..1838  
/gene="TEM8"  
/product="tumor endothelial marker 8 precursor"  
/protein\_id="AAK52094.1"  
/translaction="WATERALGIGFOMLSLATVLICAGGGRREDGCPACYGFP  
LYFLIDKSGSVLHNMNBIYYFVEQLAKFTSPQIRMSFIVFSTGTLMKLTEDREOI  
ROGLIEBQKTVLPGSDTMEGFERASQIYVENOGYRTASVIALTDGELHEDLFY  
SEBANSRDLGAIYCVGKDFNETQLARIADSKDHYFVNDGFOALQGIHLSILKK  
SCIEILAEPTICAGBSFQVVRNGFPHANVRVLCSPKINSVLINPEFVED  
TYLILPAILKEIKGKAAALQVMDNLSITSSVITTTGSDGSLIALILIFLL  
ALALIMFMPLECCVTIKEYPPPAESEEDDDGPKKMTVTNVAASYGAGVGITK  
RMEVWGEKSTEBGALKAKARVMPQOEYFEPERNLNMMNRSPSPKXYSPI  
KGDALMVLIRKGYDRSVNRPQEDTRCINFRVKNQPAKPLNNAVHTSSPP  
APVTPPPAPHCPPPPSAFPPIPSPESTLPPPPQAPPPNAPPPSPRPPPSV"



Db	353	TCCTTATGTTTCTTCCACCCGAGAAACACTTAATGAACCTACAGAAACAGAGAA	412
Qy	301	CAATCCGTCAGAGGCTTGAAGAACTCCAGAAAGTTTCGCGAGAGACACTTACATG	360
Db	413	CAATCCGTCAGAGGCTTGAAGAACTCCAGAAAGTTTCGCGAGAGACACTTACATG	472
Qy	361	CATGAGATTTGAAGAGCCGAGTACAGATTTATTAAGAAACAGAAAGGATACAG	420
Db	473	CATGAGATTTGAAGAGCCGAGTACAGATTTATTAAGAAACAGAAAGGATACAG	532
Qy	421	ACAGCAGGCTCATCTTCTTCTTGAATGAGAACTCCAGAAAGTTTCGCGAGAG	480
Db	533	ACAGCAGGCTCATCTTCTTCTTGAATGAGAACTCCAGAAAGTTTCGCGAGAG	592
Qy	481	TCAGAGAGGAGGCTTATAGGCTCGAGATCTTGGCAATTTGTTAGCTGTTGATG	540
Db	593	TCAGAGAGGAGGCTTATAGGCTCGAGATCTTGGCAATTTGTTAGCTGTTGATG	652
Qy	541	AAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGCGACAGTAAAGATCATG	600
Db	653	AAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGCGACAGTAAAGATCATG	712
Qy	601	GTGAATGACGGCTTCAGGCTTCGAGGCTCATCTCAATTTTGAAGAGTCTTCC	660
Db	713	GTGAATGACGGCTTCAGGCTTCGAGGCTCATCTCAATTTTGAAGAGTCTTCC	772
Qy	661	ATCGAATTTCTAGCAGCTGAACATCCACATATGAGAGAGTCAATTCAGTTGTC	720
Db	773	ATCGAATTTCTAGCAGCTGAACATCCACATATGAGAGAGTCAATTCAGTTGTC	832
Qy	721	GTGAGAGGAAACGGCTTCGAGCATGCGCGCAAGTGAAGAGGCTTCGAGCTTCA	780
Db	833	GTGAGAGGAAACGGCTTCGAGCATGCGCGCAAGTGAAGAGGCTTCGAGCTTCA	892
Qy	781	ATCATGACTGGTCTGACACTCAATGAGAAAGCCCTTTCTGTGGAAGACATTTAT	840
Db	893	ATCATGACTGGTCTGACACTCAATGAGAAAGCCCTTTCTGTGGAAGATTTAT	952
Qy	841	TGTCCAGGCGCTATCTTAAAGAGTTGAGTGAAGTGCACCTCAGTACAGTGAAC	900
Db	953	TGTCCAGGCGCTATCTTAAAGAGTTGAGTGAAGTGCACCTCAGTACAGTGAAC	1012
Qy	901	GATGGCCTCTCTTTATCTCCAGTCTGTGATCATCACACACACACTGT	951
Db	1013	GATGGCCTCTCTTTATCTCCAGTCTGTGATCATCACACACACACTGT	1063
RESULT 8	2112 bp mRNA linear PRI 29-JUN-2004		
LOCUS	BC012074 Homo sapiens anhrax toxin receptor 1, transcript variant 3, mRNA		
DEFINITION	(cDNA clone MGC:19967 IMAGE:4563020), complete cds.		
ACCESSION	BC012074		
VERSION	BC012074.1 GI:15082332		
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2112)		
	Klausner R.D., Collins F.S., Wagner L.H., Dergs J.G., Strausberg R.L., Feingold E.A., Grouse L.H., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marubina K., Farmer A.A., Rubin G.M., Hong L., Stepleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheer T.E., Brownstein M.J., Ueda T.B., Tohriyuki S., Carinici P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wotley K.C., Hale S., Garcia A.M., Gay L.J., Holly S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
TITLE	Pahay, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smilins, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 2112)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-AUG-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: DCTD/DRP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada <a href="mailto:info@cgsc.bc.ca">info@cgsc.bc.ca</a> Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palomquist, Anca Petrescu, Anna Liisa Prabh, Parvaneh Saeedi, Jr Santos, Angelique Scherch, Ursula Skalek, Duane Smalley, Jeff Stott, Miranda Tsai, George Yang, Jacque Schein, Asim Siddiqui, Rob Holt, Marco Marra.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAL Plate: 29 Row: 9 Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 16933552.		
SOURCE	Location/Qualifiers 1..2112 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:19967 IMAGE:4563020" /tissue_type="Kidney, renal cell adenocarcinoma" /clone_id="NIH_MGC_14" /lab_host="DH10B-R" /note="Vector: pOTB7" 1..2112 /gene="ANTXR1" /note="synonyms: FLJ10601, FLJ21776, TEN8, ATR" /db_xref="LOCUSID:84168" /db_xref="MIM:606410" 113..1114 /gene="ANTXR1" /codon_start=1 /product="tumor endothelial marker 8, isoform 3 precursor" /protein_id="AAH12074.1" /db_xref="GI:15082333" /db_xref="MIM:606410" /translation="MARRERALGIGROWSLATVVICAGGGRREDGPGACGGFD LYFLUDSGSVLHNMNITYFVEQLAKFISPOURSFIVPSTRGTLMLKTBREOT ROGIELOKVLPGDITMHEGPERASQIYENRQGRVSAIIALDGEIHEDLFFY SERLAANSRLQALIVCVGVDPAETQLARIADSKHVFVNGGFOALOGIHSILK SCIEILAAESTICAGSPVNVGNGFRHARVNDVLCSEKINDSVTLNKEPSPVED TYLCPAPILKEVGMKALQVSNMDGSLFSSSVIITTHCSLHKIASGPTAACMG"		
CDs	gene		
ORIGIN	Query Match 86.0%; Score 949.4; DB 9; Length 2112; Best local similarity 99.9%; Pred. No. 2.6e-272;		



Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	ATGGCCAGGCGGAGGAGGAGCCCTCGGCAATCGGCTTCCAGTGGCTCTCTTTGGCCACT	60
DB	113	ATGGCCAGGCGGAGGAGGAGCCCTCGGCAATCGGCTTCCAGTGGCTCTCTTTGGCCACT	172
QY	61	CTGGTCTCATCTGCGCGCGGCGAAGGAGGAGGAGATGAGGAGTCCAGCCCTGCTAC	120
DB	173	CTGGTCTCATCTGCGCGCGGCGAAGGAGGAGGAGATGAGGAGTCCAGCCCTGCTAC	232
QY	121	GCGGATTTGACCTGTACTTCTTTCAGCAAAATCAGGAAGTGTCTGACCACTGGAAAT	180
DB	233	GCGGATTTGACCTGTACTTCTTTCAGCAAAATCAGGAAGTGTCTGACCACTGGAAAT	292
QY	181	GAATATTAATCTTGTGGAAGAGTTGGCTCAAAATTCATCAGGCCCAAGTTGGAATG	240
DB	293	GAATATTAATCTTGTGGAAGAGTTGGCTCAAAATTCATCAGGCCCAAGTTGGAATG	352
QY	241	TCTTTATTTGTTTCTCCACCCGAGGAAACAACCTTATGAAACTGACAGAAACAGAA	300
DB	353	TCTTTATTTGTTTCTCCACCCGAGGAAACAACCTTATGAAACTGACAGAAACAGAA	412
QY	301	CAATTCCTGCAAGGCTTGAAGAACTCCAGAAAAGTTTCCGAGAGAGACACTTACATG	360
DB	413	CAATTCCTGCAAGGCTTGAAGAACTCCAGAAAAGTTTCCGAGAGAGACACTTACATG	472
QY	351	CATGAAGATTGAAAAGGCGCAGTAGAGAGATTTATTAAGAAAAGACAAAGGTTACAG	420
DB	473	CATGAAGATTGAAAAGGCGCAGTAGAGAGATTTATTAAGAAAAGACAAAGGTTACAG	532
QY	421	ACAGCCAGGCTCATATGCTTTGACCTGATGAGAACTCCATGAAGATCTCTTTTCTAT	480
DB	533	ACAGCCAGGCTCATATGCTTTGACCTGATGAGAACTCCATGAAGATCTCTTTTCTAT	592
QY	461	TCAGAGAGAGAGGCTTAATAGTCTGAGATCTTGTGCAATTTGTTTACTGTGTGTGTG	540
DB	593	TCAGAGAGAGAGGCTTAATAGTCTGAGATCTTGTGCAATTTGTTTACTGTGTGTGTG	652
QY	541	AAAGATTTCAATGAGACACAGCTGGCCGAGATTTGGGACAGTAAAGATCATGTGTTCC	600
DB	653	AAAGATTTCAATGAGACACAGCTGGCCGAGATTTGGGACAGTAAAGATCATGTGTTCC	712
QY	601	GTGAATGACGGCTTCAAGCTCTGCAAGGCACTCATCTCAATTTTGAAGAAGTCTGAC	660
DB	713	GTGAATGACGGCTTCAAGCTCTGCAAGGCACTCATCTCAATTTTGAAGAAGTCTGAC	772
QY	661	ATCGAATTTCTAGCAGCTGGAACCATCCAGATATGTGAGAGAGATCATTTCAATTGTC	720
DB	773	ATCGAATTTCTAGCAGCTGGAACCATCCAGATATGTGAGAGAGATCATTTCAATTGTC	832
QY	721	GTGAGAGGAAACGGCTTCCGACATGCGCGCAAGTGGACAGAGGCTCTGACAGTTCAAG	780
DB	833	GTGAGAGGAAACGGCTTCCGACATGCGCGCAAGTGGACAGAGGCTCTGACAGTTCAAG	892
QY	781	ATCAATGACTCGGTCACTCAATGAGAAACCTTTTCTGTGAAAGACACTTATTTACTG	840
DB	893	ATCAATGACTCGGTCACTCAATGAGAAACCTTTTCTGTGAAAGATCTTATTTACTG	952
QY	841	TGTCCAGGCGCTTATCTTAAAGAGTGGGATGAAAGTGCACCTCAGGTCAAGTGAAC	900
DB	953	TGTCCAGGCGCTTATCTTAAAGAGTGGGATGAAAGTGCACCTCAGGTCAAGTGAAC	1012
QY	901	GATGGCTCTCTTTTATCTCAGATTCTGATCATCATCACCAACACACTGT	951
DB	1013	GATGGCTCTCTTTTATCTCAGATTCTGATCATCATCACCAACACACTGT	1063

RESULT 9  
AX393256 5220 bp DNA linear PAT 23-MAR-2002  
LOCUS AX393256  
DEFINITION Sequence 186 from Patent W00210217.  
ACCESSION AX393256  
VERSION AX393256.1 GI:19701303

KEYWORDS			
SOURCE			
ORGANISM			
Mus musculus (house mouse)			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1..5220			
/organism="Mus musculus"			
/mol_type="unassigned DNA"			
/db_xref="taxon:10090"			
ORIGIN			
Query Match 77.3%; Score 853.4; DB 6; Length 5220;			
Best Local Similarly 87.3%; Pred. No. 1.8e-243;			
Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;			
QY	24	CCTGGCATGGGCTTCCAGTGGCTCTCTTGGCCACTCTGGTGTCAATCGCGCGGACA	83
DB	291	CCTGGATGGGCTTCCAGTGGCTCTCTTGGCCACTCTGGTGTCAATCGCGCGGACA	350
QY	84	AGGGGACGACGAGGAGATGGGGGTCCAGCTGCTACGGCGAATTTGACCTTACAT	143
DB	351	CGGGGGCGCGGAGGATGGGGGACAGCTTGTACGAGGATTTGACCTTACAT	410
QY	144	TTTGGCAATCAGAAAGTGTCTGACCACTGGAATGAATCTTATCTTTGGAAACA	203
DB	411	CCTGGAACAAGTCAAGAAAGTGTCTGACCACTGGAATGAATCTTATCTTTGGAAACA	470
QY	204	GTGGCTCACAAATTCATCAGCCACAGTTGAAATGCTTTATGTTTCTCACCCG	263
DB	471	GTGGCTCATAGATTCATCAGCCACAGTTGAAATGCTTTATGTTTCTCACCCG	530
QY	264	AGGAACAACCTTAAATGAACTGACAGAAACAGAAACAAATCCGTCAAGGCTTAAAGA	323
DB	531	AGGAACAACCTTAAATGAACTGACAGAAACAGAAACAAATCCGTCAAGGCTTAAAGA	590
QY	324	ACTCCAGAAAGTTTCTGCCAGAGAGACCTTACATGATGAAAGATTTGAAGGCGCAG	383
DB	591	GCTCCAGAAAGTTTCTGCCAGAGAGACCTTACATGATGAAAGATTTGAAGGCGCAG	650
QY	384	TGAGCAGATTATATGAAACAGACAAGGGTACAGGACAGCCAGCTCATCTGCTT	443
DB	651	TGAGCAGATTATATGAAACAGACAAGGGTACAGGACAGCCAGCTCATCTGCTT	710
QY	444	GACTGATGAGAACTCCATGAAAGATCTTTTCTATTACAGAGGAGGCTTAAATAGTC	503
DB	711	GACGGATGGGAGCTGACAGAGACCTCTTCTTACTCAGAGGAGGCTTAAACGATC	770
QY	504	TCGAGATTTGTGTGCAATTTGTTTACTGTGTGTGTGTAAGATTTTCATGACACAGCT	563
DB	771	CCGAGACCTTGTGTGCAATTTGTTTACTGTGTGTGTGTAAGATTTTCATGACACAGCT	830
QY	564	GAGCCGAGTTGCGGACAGTAAAGATCAATGTGTTTCCGTAATGACAGGCTTTCAGGCTCT	623
DB	831	GAGCCGAGTTGCGGACAGTAAAGATCAATGTGTTTCCGTAATGACAGGCTTTCAGGCTCT	890
QY	624	GCAAGCATCATCACTCAATTTTGAAGAGTCTGCAATGAAATTTCTAGAGCTGAAC	683
DB	891	CCAAGCATTATCACTCAATTTTGAAGAGTCTGCAATGAAATTTCTAGAGCTGAAC	950
QY	684	ATCCACCATATGTGAGAGAGATCATTTCAAGTTGTGTGAGAGAAAGGCTTCCGACA	743
DB	951	ATCCACCATATGTGAGAGAGATCATTTCAAGTTGTGTGAGAGAAAGGCTTCCGACA	1010
QY	744	TGCCGCAAGTGGACAGGGTCTCTGAGCTTCAAGTCAATGATCGTCACTCA	803
DB	1011	TGCCGCAATGTGGACAGGGTCTCTGAGCTTCAAGTCAATGATCGTCACTCA	1070



QY 804 TGAGAAAGCCCTTTCTGTGGAGAGACCTTATTACTGTGTCCAGCCCTATCTTAAAGA 863  
 DB 1071 TGAGAAAGCCCTTTCTGTGGAGAGACCTTATTACTGTGTCCAGCCCTATCTTAAAGA 1130  
 QY 864 AGTTGGCAATGAAGCTGCACTCCAGGTGAGCATGAAAGCATGGCTCTCTTTTATCTCCAG 923  
 DB 1131 AGTTGGCAATGAAGCTGCACTCCAGGTGAGCATGAAAGCATGGCTCTCTTTTATCTCCAG 1190  
 QY 924 TTCTGTCAATCATCACCAACACACACTGTCTGACGGTTCATCTGGCCATCGCCCTGCT 983  
 DB 1191 TTCTGTCAATCATCACCAACACACACTGTCTGACGGTTCATCTGGCCATCTGCTGCT 1250  
 QY 984 GATCTCTTCTGTCTGTCTGAGCCCTGCTCTCTGTGGTTCGTGCCCCCTGTGTGAC 1043  
 DB 1251 GGTCTCTCTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1310  
 QY 1044 TGTGATTTATCAAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
 DB 1311 AGTGAATCATCAAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361

RESULT 10  
 AX393370  
 LOCUS AX393370 5220 bp DNA linear PAT 23-MAR-2002  
 DEFINITION Sequence 300 from Patent WO0210217.  
 ACCESSION AX393370  
 VERSION AX393370.1 GI:19701356  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 St Croix, B., Kinzler, K.W., and Vogelstein, B.  
 Endothelial cell expression patterns  
 Patent: WO 0210217-A 300 07-FEB-2002;  
 The Johns Hopkins University (US)  
 Location/Qualifiers  
 1..5220  
 /organism="Mus musculus"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:10090"

ORIGIN  
 Query Match 77.3% Score 853.4; DB 6; Length 5220;  
 Best Local Similarity 87.3%; Pred. No. 1.8e-243;  
 Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

DB 651 TGAGCAAGATTTACTATGAGAACAGTCAAGGATACAGAGACGGCGACGTCATCATCGCGTT 710  
 QY 444 GACGTATGAGAGACCTCCATGAAATCTCTTTTCTATTCAGAGAGAGGCGCTAATATGTC 503  
 DB 711 GACGATGGAGAGCTGACAGAGACCTCTCTTCTACTAGAGAGAGGCTTAACCGATC 770  
 QY 504 TCGAGATCTGGTGCATTTGTTTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 563  
 DB 771 CCGAGACCTTGGTGGATTTGTTTACTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 830  
 QY 564 GGCCTGGATTCGAGACAGTAAATCATGTTTCCGTTAATGACGCTTTCAGGCTCT 623  
 DB 831 GGTCTGGAATTCAGACAGTAAAGACAGTGTTCCTGTGAACGACGCTTTCAGGCTCT 890  
 QY 624 GCAAGGCAATTCACCTCAATTTTGAAGAAGCTCCGATCCGAATTTCAAGAGCTGAAC 683  
 DB 891 CCAAGGCAATTCACCTCAATTTTGAAGAAGCTCCGATCCGAATTTCAAGAGCTGAAC 950  
 QY 684 ATCCACATATGTCAGAGAGTCAATTCAGTTGCTGAGAGAGAAACGGCTTCCGACA 743  
 DB 951 ATCCACATATGTCAGAGAGTCAATTCAGTTGCTGAGAGAGAAATGGCTTCCGACA 1010  
 QY 744 TGCCCGCAACGTGACAGAGGTCCTCTGACGCTTCAAGATCAATGACTGCTACACTGAA 803  
 DB 1011 TGCCCGCAATGTCAGAGAGGTCCTCTGACGCTTCAAGATCAATGACTGCTACACTGAA 1070  
 QY 804 TGAGAAAGCCCTTTCTGTGGAGAGACCTTATTACTGTGTCCAGGCTTATCTTAAAGA 863  
 DB 1071 TGAGAAAGCCCTTTCTGTGGAGAGACCTTATTACTGTGTCCAGGCTTATCTTAAAGA 1130  
 QY 864 AGTTGGCAATGAAGCTGCACTCCAGGTGAGCATGAAAGCATGGCTCTCTTTTATCTCCAG 923  
 DB 1131 AGTTGGCAATGAAGCTGCACTCCAGGTGAGCATGAAAGCATGGCTCTCTTTTATCTCCAG 1190  
 QY 924 TTCTGTCAATCATCACCAACACACACTGTCTGACGGTTCATCTGGCCATCGCCCTGCT 983  
 DB 1191 TTCTGTCAATCATCACCAACACACACTGTCTGACGGTTCATCTGGCCATCTGCTGCT 1250  
 QY 984 GATCTCTTCTGTCTGTCTGAGCCCTGCTCTCTGTGGTTCGTGCCCCCTGTGTGAC 1043  
 DB 1251 GGTCTCTCTCTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1310  
 QY 1044 TGTGATTTATCAAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1094  
 DB 1311 AGTGAATCATCAAGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1361

RESULT 11  
 AF378762  
 LOCUS AF378762 5220 bp mRNA linear ROD 07-OCT-2001  
 DEFINITION Mus musculus tumor endothelial marker 8 precursor (Tem8) mRNA,  
 complete cds.  
 ACCESSION AF378762  
 VERSION AF378762.1 GI:15987504  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Kinzler, K.W., and St. Croix, B.  
 Cell surface tumor endothelial markers are conserved in mice and  
 humans  
 Cancer Res. 61 (18), 6649-6655 (2001)  
 MEDLINE 21443268  
 PUBMED 11559528  
 2 (bases 1 to 5220)  
 Carson-Walter, E.B., Vogelstein, B., Kinzler, K.W., and St. Croix, B.  
 Direct Submission  
 Submitted (10-MAY-2001) Oncology, Johns Hopkins University, 1650  
 Orleans Street, Baltimore, MD 21231, USA



Qy	181	GAATCTATCTACCTTTGTGGAAACAGTTGGCTCCAAATTCATCATAGCCACAGTTTGAATG	240
Db	489	GAATCTATCTACCTTTGTGGAAACAGTTGGCTCCAAATTCATCATAGCCACAGTTTGAATG	548
Qy	241	TCCTTTATGTTTCTTCACCCGAGGAAACACTTAATGAAACTGACAGAAACAGAGA	300
Db	549	TCCTTTATGTTTCTTCACCCGAGGAAACACTTAATGAAACTGACAGAAACAGAGA	608
Qy	301	CAAAATCCGTCMAAGGCTTAGAAGATCTCCAGAAAGTTCTGCCAGAGAGACCTTAATG	360
Db	609	CAAAATCCGTCMAAGGCTTAGAAGATCTCCAGAAAGTTCTGCCAGAGAGACCTTAATG	668
Qy	361	CATGAGAAGATTGAAAGAGGCGCAGTGGAGCATTTATTAAGAAAACAGACAAGGGTACAG	420
Db	669	CATGAGAAGATTGAAAGAGGCGCAGTGGAGCATTTATTAAGAAAACAGACAAGGGTACAG	728
Qy	421	ACAGCCAGCGTCATCATCTTGTGACTGATGAGAACTCCATGAAGATCTCTTTTCTAT	480
Db	729	ACAGCCAGCGTCATCATCTTGTGACTGATGAGAACTCCATGAAGATCTCTTTTCTAT	788
Qy	481	TCAGAGAAGGAGGCTTAATAGTCTCGAATCTTTGTGCAATTTGTTACTGTGTGTGTG	540
Db	789	TCAGAGAAGGAGGCTTAATAGTCTCGAATCTTTGTGCAATTTGTTACTGTGTGTGTG	848
Qy	541	AAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGCGGACAGTAAGGATCATGTGTTTCCC	600
Db	849	AAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGCGGACAGTAAGGATCATGTGTTTCCC	908
Qy	601	GTAATGACGCGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAAGTCTGCG	660
Db	909	GTAATGACGCGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAAGTCTGCG	968
Qy	661	ATCGAAATTCATGACAGCTGGAACCATTCACCAATGTGCGAGAGAGTCAATTTCAAATTTGTC	720
Db	969	ATCGAAATTCATGACAGCTGGAACCATTCACCAATGTGCGAGAGAGTCAATTTCAAATTTGTC	1028
Qy	721	GTCGAGAGGAAACGCGCTTCGCAATGCCGCAACGTGACAGAGGTCCTCTGACGCTTCAAG	780
Db	1029	GTCGAGAGGAAACGCGCTTCGCAATGCCGCAACGTGACAGAGGTCCTCTCTGACGCTTCAAG	1088
Qy	781	ATCAATGACTCGGTCACTCACTCAATGAG	807
Db	1089	ATCAATGACTCGGTCACTCACTCAATGAG	1115
RESULT 13			
LOCUS	CQ498433	1718 bp	DNA
DEFINITION	Sequence 30300 from Patent WO0160860.		
ACCESSION	CQ498433		
VERSION	CQ498433.1	GI:41464069	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1	Schlegel,R., Endege,W.O. and Monahan,J.E.	
AUTHORS		Genes differentially expressed in human prostate cancer and their	
TITLE		use	
JOURNAL		Patent: WO 0160860-A 30300 23-AUG-2001;	
FEATURES		Millennium Predictive Medicine, Inc. (US)	
source		Location/Qualifiers	
		1..1718	
		/organism="Homo sapiens"	
		/mol_type="unassigned DNA"	
		/db_xref="taxon:9606"	
ORIGIN			
Query Match	72.7%;	Score 802.2;	DB 6; Length 1718;
Best Local Similarity	99.6%;	Pred. No.36-228;	
Matched 804;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;

QY	1	ATGGCGCAGCGCGGAGCGGAGAGCGCCCTCGGCATCTCGGCTTCAGATGGCTCTCTTTGGCCACT	60
Db	413	ATGGCGCAGCGCGGAGCGGAGAGCGCCCTCGGCATCTCGGCTTCAGATGGCTCTCTTTGGCCACT	472
QY	61	CTGTGTCTCATCTGGCGCGGCGAAGGGGAGACGACGAGAGAGATGGGGGTCCAGCTCTGAC	120
Db	473	CTGTGTCTCATCTGGCGCGGCGAAGGGGAGACGACGAGAGAGATGGGGGTCTCAGCTCTGAC	532
QY	121	GGCGGATTTGACCTGTACTTCAATTTTGGACAATCAGAGAGTGTGCTGCACACTGGAAAT	180
Db	533	GGCGGATTTGACCTGTACTTCAATTTTGGACAATCAGAGAGTGTGCTGCACACTGGAAAT	592
QY	181	GAAATCTATTACTTTGTGGAAACAGTGGCTCAAAATTCATCAGCCCAAGTTGGAGATG	240
Db	593	GAAATCTATTACTTTGTGGAAACAGTGGCTCAAAATTCATCAGCCCAAGTTGGAGATG	652
QY	241	TCTTTATATGTTTCTCCACCGGAGGAAACAACCTTAATGAAACTGACAGAAACAGAGAA	300
Db	653	TCTTTATATGTTTCTCCACCGGAGGAAACAACCTTAATGAAACTGACAGAAACAGAGAA	712
QY	301	CAATTCGGTCAAAGCCTTGAAAGACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG	360
Db	713	CAATTCGGTCAAAGCCTTGAAAGACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG	772
QY	361	CATGAAGATTTGAAAGGGCCAGTAGAGATTTATTTGAAACACAGCAAGGGATCAGG	420
Db	773	CATGAAGATTTGAAAGGGCCAGTAGAGATTTATTTGAAACACAGCAAGGGATCAGG	832
QY	421	ACAGCCAGCGTCATCATGCTTGTGAAGTATGAGAGAACTCCATGAAGATCTCTTTCTAT	480
Db	833	ACAGCTAGCGTCATCATGCTTGTGAAGTATGAGAGAACTCCATGAAGATCTCTTTCTAT	892
QY	481	TCAGAGAGGAGGCTAATAGTCTTCGAGATCTTGGTGCATTTGTTTACTGTGTGGTGTG	540
Db	893	TCAGAGAGGAGGCTAATAGTCTTCGAGATCTTGGTGCATTTGTTTACTGTGTGGTGTG	952
QY	541	AAAGATTTCAATGAGACACAGCTGGCCGAGATGGCGGACAGTAAAGGATCATGTGTTTCCC	600
Db	953	AAAGATTTCAATGAGACACAGCTGGCCGAGATGGCGGACAGTAAAGGATCATGTGTTTCCC	1012
QY	601	GTTGAATGACGGCTTCAGGCTCTGCGAAGGACATCCACTCAATTTTGAAGAGTCCCTGC	660
Db	1013	GTTGAATGACGGCTTCAGGCTCTGCGAAGGACATCCACTCAATTTTGAAGAGTCCCTGC	1072
QY	661	ATCGAAATTTAGACAGCTGAAACCATCCACATATGTGACAGAGAGTCAATTTCAAGTTGTC	720
Db	1073	ATCGAAATTTAGACAGCTGAAACCATCCACATATGTGACAGAGAGTCAATTTCAAGTTGTC	1132
QY	721	GTTGAGAGAAAACGGCTTCGACATGCGCCGCAACGTCGACAGGGTCTCTGCAGCTTCAAG	780
Db	1133	GTTGAGAGAAAACGGCTTCGACATGCGCCGCAACGTCGACAGGGTCTCTGCAGCTTCAAG	1192
QY	781	ATCATATGACTCGTCCACTCATATGAG 807	
Db	1193	ATCATATGACTCGTCCACTCATATGAG 1219	
RESULT 14			
LOCUS	BD156323	1436 bp	DNA linear PAT 17-JAN-2003
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD156323		
VERSION	BD156323.1	GI:27862081	
KEYWORDS	JP 2002191363-A/11166.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Ota,T., Ito,T., Nishikawa,T., Hayashi,K., Saio,K., Yamamoto,J., Ishii,S., Sugiyama,T., Wakematsu,A., Nagai,K. and Otsuki,T.		
TITLE	Primer for synthesizing full-length cDNA and use thereof		

JOURNAL Patent: JP 2002191363-A 11166 09-JUL-2002;  
HELIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002191363-A/11166  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 2000280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
PI SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC  
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers  
FT CDS Location/Qualifiers (380)..(1033).  
FEATURES  
Source Location/Qualifiers  
1..1436  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 71.3%; Score 787; DB 6; Length 1436;  
Best Local Similarity 99.3%; Pred. No. 1e-223;  
Matches 801; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
1 ATGGCCACGGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTTTGGCCACT 60  
144 ATGGCCACGGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTC-ACGGCCACT 202  
61 CTGGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGTCCAGCTGTCTAC 120  
203 CTGGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGTCCAGCTGTCTAC 262  
121 GCGGATTTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTCTGACCACTGTGAAT 180  
263 GCGGATTTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTCTGACCACTGTGAAT 322  
181 GAAATCTATCTTTGTGGAAAGATTGGTCTCAAAATTCATGAGCCCAAGTTGGAATG 240  
323 GAAATCTATCTTTGTGGAAAGATTGGTCTCAAAATTCATGAGCCCAAGTTGGAATG 382  
241 TCCCTTATTTGTTTCTCCACCCGAGAAACAACTTAATGAACTGACAGAAACAGAGAA 300  
383 TCCCTTATTTGTTTCTCCACCCGAGAAACAACTTAATGAACTGACAGAAACAGAGAA 442  
301 CAATCCGTCAGAGGCTTAGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 360  
443 CAATCCGTCAGAGGCTTAGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 502  
361 CATGAAGATTTGAAAGGCGGAGAGAGATTTATATGAAAACAGACAAAGGTTACAGG 420  
503 CATGAAGATTTGAAAGGCGGAGAGAGATTTATGAAAACAGACAAAGGTTACAGG 562  
421 ACAGCAGAGCTCATCTTGTCTTGAAGTGAAGAACTCCATGAAGATCTCTTTTCTAT 480  
563 ACAGCAGAGCTCATCTTGTCTTGAAGTGAAGAACTCCATGAAGATCTCTTTTCTAT 622  
481 TCAGAGAGGAGAGCTTAATAGTCTCGAGATCTTGTGCAATGTTTACTGTGTGTGTG 540  
623 TCAGAGAGGAGAGCTTAATAGTCTCGAGATCTTGTGTGCAATGTTTACTGTGTGTGTG 682  
541 AAAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGGAGACGTAAGATCATGTGTTTCCC 600  
683 AAAAGATTTCAATGAGACACAGCTGGCCCGGATTTGGGAGACGTAAGATCATGTGTTTCCC 742  
601 GTGAATGAGAGGCTTTCAGGCTCTGCAAGGATCATCTCAATTTTGAAGAGTCTGCG 660  
743 GTGAATGAGAGGCTTTCAGGCTCTGCAAGGATCATCTCAATTTTGAAGAGTCTGCG 802  
661 ATCGAAATTTCTAGCAGCTGAGACATCCACATATGTGAGAGAGAGTCAATTTCAAGTTGTC 720

|||||  
Db 803 ATCGAAATTTCTAGCAGCTGAGACCATCCATATGTGACAGAGAGTCAATTTCAAGTTGTC 862  
Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTTGACAGGGTCTCTGACGTTCAAG 780  
Db 863 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTTGACAGGGTCTCTGACGTTCAAG 922  
Qy 781 ATCAATGACTCGGTCACTCACTCAATGAG 807  
Db 923 ATCAATGACTCGGTCACTCACTCAATGAG 949  
RESULT 15  
AX458374 1436 bp DNA linear PAT 08-JUL-2002  
LOCUS Sequence 9 from Patent WO0246228.  
DEFINITION AX458374  
ACCESSION AX458374  
VERSION AX458374.1 GI:21725044  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 Young, J.A., Bradley, K.A., Collier, R.J. and Morigide, J.S.  
TITLE Receptor for b. Anthracis toxin  
JOURNAL Patent: WO 0246228-A 9 13-JUN-2002;  
WISCONSIN ALUMNI RESEARCH FOUNDATION (US)  
FEATURES  
Source Location/Qualifiers  
1..1436  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
380..1036  
/note="unamed protein product"  
/codon\_start=1  
/protein\_id="CAD38352.1"  
/db\_xref="GI:21725045"  
/translators="MSPIVSTRGTTLMKLTEDREQIROGLEBLQKYLPGSDTYMHEG  
FERASEOIVYENRQGYRTASVIALITDGELEHEDLPFYSEBEANRSDLAIVYCVGV  
DFNETOLARIADSKDHYFPVNDGFOALOGIITHSLTKSCIEIILAAESTICAGSFQY  
VVRGNSFRHARNDRVLCFRKINDSVLSLSLQSPWVSISGFRGNSHPLCLPARPHT  
"  
ORIGIN  
Query Match 71.3%; Score 787; DB 6; Length 1436;  
Best Local Similarity 99.3%; Pred. No. 1e-223;  
Matches 801; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
1 ATGGCCACGGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTTTGGCCACT 60  
Db 144 ATGGCCACGGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTC-ACGGCCACT 202  
Qy 61 CTGGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGTCCAGCTGTCTAC 120  
Db 203 CTGGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGTCCAGCTGTCTAC 262  
Qy 121 GCGGATTTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTCTGACCACTGTGAAT 180  
Db 263 GCGGATTTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTCTGACCACTGTGAAT 322  
181 GAAATCTATCTTTGTGGAAAGATTGGTCTCAAAATTCATGAGCCCAAGTTGGAATG 240  
Db 323 GAAATCTATCTTTGTGGAAAGATTGGTCTCAAAATTCATGAGCCCAAGTTGGAATG 382  
Qy 241 TCCCTTATTTGTTTCTCCACCCGAGAAACAACTTAATGAACTGACAGAGAGAGAGAA 300  
Db 383 TCCCTTATTTGTTTCTCCACCCGAGAAACAACTTAATGAACTGACAGAGAGAGAGAA 442  
Qy 301 CAATCCGTCAGAGGCTTAGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 360  
Db 443 CAATCCGTCAGAGGCTTAGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 502

```
Qy 361 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTATGAAAACAGACAAAGGTAACAG 420
    |||||
Db 503 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTATGAAAACAGACAAAGGTAACAG 562
    |||||
Qy 421 ACAGCAGCGGTATCATTCCTTTGACTGTGAGAACTCCATGAAGATCTCTTTTCTAT 480
    |||||
Db 563 ACAGCTAAGGTATCATTCCTTTGACTGTGAGAACTCCATGAAGATCTCTTTTCTAT 622
    |||||
Qy 481 TCAGAGAGGGAGCTTAATAGTCTCGAGATCTTGAGCAATTGTTTACTGTGTGTG 540
    |||||
Db 623 TCAGAGAGGGAGCTTAATAGTCTCGAGATCTTGAGCAATTGTTTACTGTGTGTG 682
    |||||
Qy 541 AAAGATTTCAATGAGACACAGCTGGCCCGAATTCGGAAGATCATGTGTTCC 600
    |||||
Db 683 AAAGATTTCAATGAGACACAGCTGGCCCGAATTCGGAAGATCATGTGTTCC 742
    |||||
Qy 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAAGTCTGC 660
    |||||
Db 743 GTGAATGACGGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAAGTCTGC 802
    |||||
Qy 661 ATCGAAATTTCTAGCAGCTGAACCATCCATATATGTGAGAGAGTCAATTCAGTTGTC 720
    |||||
Db 803 ATCGAAATTTCTAGCAGCTGAACCATCCATATATGTGAGAGAGTCAATTCAGTTGTC 862
    |||||
Qy 721 GTGAGAGGAACGGCTTCCGACATGCCGCAAGTGAACGGGCTCTGACAGTTCAAG 780
    |||||
Db 863 GTGAGAGGAACGGCTTCCGACATGCCGCAAGTGAACGGGCTCTGACAGTTCAAG 922
    |||||
Qy 781 ATCAATGACTCGGTCACTCAATGAG 807
    |||||
Db 923 ATCAATGACTCGGTCACTCAATGAG 949
    |||||
```

Search completed: June 14, 2005, 15:32:17  
Job time : 5150 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: June 14, 2005, 11:26:37 ; Search time 700 Seconds  
(without alignments)  
9336.278 Million cell updates/sec

Title: US-09-970-076-1\_COPY\_104\_1207

Perfect score: 1104  
Sequence: 1 atggtcacgagcggagcggag.....gtgaggaataataataaaa 1104

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_GeneSeq\_16Dec04:\*

1: geneSeq19808:\*\n2: geneSeq19908:\*\n3: geneSeq20008:\*\n4: geneSeq20018:\*\n5: geneSeq20028:\*\n6: geneSeq20038:\*\n7: geneSeq20048:\*\n8: geneSeq20058:\*\n9: geneSeq20068:\*\n10: geneSeq20078:\*\n11: geneSeq20088:\*\n12: geneSeq20098:\*\n13: geneSeq20108:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1104	100.0	1414	6	ABV73881 Human ant
2	1090.8	98.8	5540	6	ABL92078 Human Tum
3	1090.8	98.8	5540	6	ABL92104 Human Tum
4	1090.8	98.8	5540	6	ABV73882 Human ant
5	1090.8	98.8	5540	10	ABX72029 DNA encod
6	1090.8	98.8	5540	10	ABX72003 DNA encod
7	1090.8	98.8	5540	13	ADR48215 Human Tum
8	1090.6	98.8	2447	4	AD05303 Human sec
9	1089.4	98.7	2086	4	AD05334 Human sec
10	960.8	87.0	1674	10	AD100549 Human TAN
11	960.8	87.0	1674	11	ADM64575 Mouse TAN
12	952.4	86.3	1650	10	AD100545 Human TAN
13	952.4	86.3	1650	10	AD100547 Human TAN
14	952.4	86.3	1650	11	ADM64589 Human TAN
15	952.4	86.3	1650	11	ADM64591 Human TAN
16	950.8	86.1	1650	10	AD100541 Human TAN
17	950.8	86.1	1650	11	ADM64585 Human TAN
18	949.4	86.0	1056	10	AD100555 Human TAN
19	949.4	86.0	1056	11	ADM64581 Human TAN
20	949.4	86.0	1713	10	AD100551 Human TAN

21	949.4	86.0	1713	11	ADM64577 Human TAN
22	949.4	86.0	2112	6	ABV73883 Human ant
23	949.4	86.0	2272	3	AAA47455 Human TAN
24	949.4	86.0	2272	10	AD100533 Human TAN
25	949.4	86.0	2272	11	AD100537 Human TAN
26	949.4	86.0	2353	11	ACN88807 Breast ca
27	949.4	86.0	1008	10	AD100557 Human TAN
28	949.4	86.0	1008	11	ADM64583 Human TAN
29	912.4	82.6	1047	10	AD100553 Human TAN
30	912.4	82.6	1047	11	ADM64579 Human TAN
31	892.6	80.9	1608	10	AD100567 Human TAN
32	867.6	78.6	1623	10	AD100543 Human TAN
33	867.6	78.6	1623	11	ADM64587 Human TAN
34	853.4	77.3	5220	6	AB192085 Mouse Tum
35	853.4	77.3	5220	6	AB192138 Mouse Tum
36	853.4	77.3	5220	10	ABX72010 DNA encod
37	853.4	77.3	5220	10	ABX72063 DNA encod
38	819	74.2	1534	10	AD100565 Human TAN
39	803.8	72.8	1402	4	AA159918 Human pol
40	803.8	72.8	1609	4	AA158132 Human pol
41	803.8	72.8	1609	5	AD098338 DNA encod
42	803.8	72.8	1609	9	ADB48098 Novel hum
43	802.2	72.7	1718	5	ABV30282 Human pro
44	787	71.3	1436	4	AAH14331 Human CDN
45	787	71.3	1436	6	ABV73884 Human ant

## ALIGNMENTS

RESULT 1	ABV73881	standard; cDNA; 1414 BP.
ID	ABV73881	standard; cDNA; 1414 BP.
XX	ABV73881	
AC	ABV73881	
XX	08-JUN-2003	(first entry)
DT	XX	
DE	Human anthrax toxin receptor cDNA.	
XX	XX	
KW	Anthrax; toxin; receptor; human; antibacterial; gene; ss.	
OS	Homo sapiens.	
XX	XX	
FT	Key	Location/Qualifiers
FT	CDS	104..1210
FT		/*tag= a
FT	sig_peptide	/product= "anthrax toxin receptor"
FT		/*tag= b
FT	mat_peptide	183..1207
FT		/*tag= c
XX	XX	
FN	W0200246228-A2.	
XX	XX	
PD	13-JUN-2002.	
XX	XX	
PF	03-OCT-2001; 2001WO-US030941.	
XX	XX	
PR	05-DEC-2000; 2000US-0251481P.	
XX	XX	
PA	(WISC) WISCONSIN ALUMNI RES FOUND.	
XX	XX	
PI	Young JAT, Bradley KA, Collier RJ, Mogridge JS;	
XX	XX	
DR	WPI; 2002-713235/77.	
DR	P-PSDB; ABP54903.	
XX	XX	
PT	Novel isolated polypeptide useful for identifying agent that prevents or	
PT	reduces effect of anthrax toxin on host cell, for treating human or non-	
PT	human animal suffering from anthrax.	
XX	XX	
PS	Claim 7; Page 27-29; 45pp; English.	





CC They are useful for inhibiting tumour growth, neoangiogenesis in subjects  
 CC bearing a vascularised tumour, polycystic kidney disease, diabetic  
 CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM  
 CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)  
 CC are disclosed, as are marker oligonucleotide sequences: tumour  
 CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal  
 CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers  
 CC (PEM) ABL91903-ABL91995

Sequence 5540 BP; 1601 A; 1306 C; 1187 G; 1446 T; 0 U; 0 Other;

Query Match 98.8%; Score 1090.8; DB 6; Length 5540;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGGCCAGGCGGCGAGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCACT 60  
 144 ATGGCCAGGCGGCGAGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCACT 203  
 61 CTGGTCTCATCTGCGCGCGGCGAGGAGGAGGAGATGGGGTCCAGCCTGCTAC 120  
 204 CTGGTCTCATCTGCGCGCGGCGAGGAGGAGGAGATGGGGTCCAGCCTGCTAC 263  
 121 GCGGATTTGACCTGTACTTCTTTTGGACAAATCAGAGAGTGTCTGACCACTGGAAT 180  
 264 GCGGATTTGACCTGTACTTCTTTTGGACAAATCAGAGAGTGTCTGACCACTGGAAT 323  
 181 GAAATCTATTCTTTGTGGAACAGTTGGCTCAGAAATTCATCAGGCCACAGTTGGAATG 240  
 324 GAAATCTATTCTTTGTGGAACAGTTGGCTCAGAAATTCATCAGGCCACAGTTGGAATG 383  
 241 TCCTTATTGTTTCTCCAGCCGAGAAACAACCTTATGAACTGACAGAAAGACAGAA 300  
 384 TCCTTATTGTTTCTCCAGCCGAGAAACAACCTTATGAACTGACAGAAAGACAGAA 443  
 301 CAAATCCGTCAAGGCGCTTGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACCTTACATG 360  
 444 CAAATCCGTCAAGGCGCTTGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACCTTACATG 503  
 361 CATGAGGATTTGAAAGGCGCGAGAGAGATTTATTTGAAACAGACAAAGGATACAGG 420  
 504 CATGAGGATTTGAAAGGCGCGAGAGAGATTTATTTGAAACAGACAAAGGATACAGG 563  
 421 ACAGCAGCGCTCATCTGCTTGAATGAGAGAACTCAGTGAAGATCTCTTTTCTAT 480  
 564 ACAGCAGCGCTCATCTGCTTGAATGAGAGAACTCAGTGAAGATCTCTTTTCTAT 623  
 481 TCAGAGAGGAGGCTTAATAGTCTCGAATCTTGGTCAATTTTACTGTGTGTG 540  
 624 TCAGAGAGGAGGCTTAATAGTCTCGAATCTTGGTCAATTTTACTGTGTGTG 683  
 541 AAAAGTTTCATGAGACACAGCTGCGCGGAGATGGGACAGTAAAGATCATGTGTTTCC 600  
 684 AAAAGTTTCATGAGACACAGCTGCGCGGAGATGGGACAGTAAAGATCATGTGTTTCC 743  
 601 GTGATGACGCGCTTCAAGCTCTCGAAGGCAATCATCACTCAATTTTGAAGAAGTCTGC 660  
 744 GTGATGACGCGCTTCAAGCTCTCGAAGGCAATCATCACTCAATTTTGAAGAAGTCTGC 803  
 661 ATCGAATTTTGAAGCAGTGAACCATCAACATATGTGAGAGAGTCAATTTCAAGTGTG 720  
 804 ATCGAATTTTGAAGCAGTGAACCATCAACATATGTGAGAGAGTCAATTTCAAGTGTG 863  
 721 GTGAGAGGAAACGCTTCCGACATGCGCGGAAACGTGACAGGCTCTCTGACGCTTCAAG 780  
 864 GTGAGAGGAAACGCTTCCGACATGCGCGGAAACGTGACAGGCTCTCTGACGCTTCAAG 923  
 781 ATCAATGATCGGTGACACTCAATGAGAAAGCCCTTTCTGTGGAAGACATTAATTACTG 840  
 924 ATCAATGATCGGTGACACTCAATGAGAAAGCCCTTTCTGTGGAAGACATTAATTACTG 983  
 841 TGTCCAGCGCTATTTTAAAGAAAGTGGCATGAAAGCTGCACTCAGGTCAAGCATGAAC 900

DB 984 TGTCCAGCGCCTATCTTAAAGAAAGTTGGCATGAAAGCTGCATCCAGGTCAAGTGAAC 1043  
 QY 901 GATGCGCTCTCTTTATCTCCAGTTCTGTCAATCATCAACCAACACACTGTTTGAAGGT 960  
 DB 1044 GATGCGCTCTCTTTATCTCCAGTTCTGTCAATCATCAACCAACACACTGTTTGAAGGT 1103  
 QY 961 TCCATCTGCGCATGCGCCTCTGTATCTGTGTTCTGTCTTACGCCCTGCTCTCTG 1020  
 DB 1104 TCCATCTGCGCATGCGCCTCTGTATCTGTGTTCTGTCTTACGCCCTGCTCTCTG 1163  
 QY 1021 TGTCTTGGGCGGCTCTGTGACATGTATTAAGAGAGGTCCTTCCACCCCTGCGAG 1080  
 DB 1164 TGTCTTGGGCGGCTCTGTGACATGTATTAAGAGAGGTCCTTCCACCCCTGCGAG 1223  
 QY 1081 GAGAGTGAAGAAA 1094  
 DB 1224 GAGAGTGAAGAAA 1237

RESULT 3  
 ABL92104  
 ID ABL92104 standard; cDNA; 5540 BP.  
 XX  
 AC ABL92104;  
 XX  
 DT 30-MAY-2002 (first entry)  
 XX  
 DE Human Tumour Endothelial Marker polynucleotide SEQ ID NO 231.  
 XX  
 KW Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cyclostatic;  
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;  
 KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;  
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;  
 KW psoriasis; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200210217-A2.  
 XX  
 PD 07-FEB-2002.  
 XX  
 PF 01-AUG-2001; 2001WO-US024031.  
 XX  
 PR 02-AUG-2000; 2000US-0222599P.  
 PR 11-AUG-2000; 2000US-0224360P.  
 PR 11-APR-2001; 2001US-0282850P.  
 XX  
 PA (UWJO ) UNIV JOHNS HOPKINS.  
 XX  
 PI St Croix B, Kinzler KW, Vogelstein B;  
 XX  
 DR WPI; 2002-291856/33.  
 DR P-PSDB; ABB90750.  
 XX  
 PT An isolated molecule comprising an antibody variable region which  
 PT specifically binds to an extracellular domain of a tumor endothelial  
 PT marker (TEM) protein, useful for inhibiting tumor growth.  
 XX  
 PS Claim 30; Page 207-209; 33pp; English.

The invention relates to an isolated molecule comprising an antibody  
 variable region which specifically binds to an extracellular domain of a  
 tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,  
 ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM  
 proteins have cyclostatic, immunostimulant and antiangiogenic activity.  
 They are useful for inhibiting tumour growth, neoangiogenesis in subjects  
 bearing a vascularised tumour, polycystic kidney disease, diabetic  
 CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM  
 CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)  
 CC are disclosed, as are marker oligonucleotide sequences: tumour  
 CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal  
 CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers  
 CC (PEM) ABL91903-ABL91995

```

XX Sequence 5540 BP; 1601 A; 1306 C; 1187 G; 1446 T; 0 U; 0 Other;
SQ
Query Match      98.8%; Score 1090.8; DB 6; Length 5540;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCAGCGGCGAGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTCTTGGCCACT 60
DB 144 ATGGCCAGCGGCGAGAGAGCCCTCGGCATCGGCTTCAGTGGCTCTCTTGGCCACT 203
QY 61 CTGGTCTCATCTGCGCGGCGAGGCGAGGAGATGGGGGTTCAGGCTGCTAC 120
DB 204 CTGGTCTCATCTGCGCGGCGAGGCGAGGAGATGGGGGTTCAGGCTGCTAC 263
QY 121 GGGCGATTGACCTGTACTTCACTTTTGGACAAATAGGAAAGTGTCTGACACACTGGAAT 180
DB 264 GGGCGATTGACCTGTACTTCACTTTTGGACAAATAGGAAAGTGTCTGACACACTGGAAT 323
QY 181 GAAATCTATTAATTGTGGAAAGTTGGCTCACAATTCATCAGCCCAAGTTGAAATG 240
DB 324 GAAATCTATTAATTGTGGAAAGTTGGCTCACAATTCATCAGCCCAAGTTGAAATG 383
QY 241 TCCTTTATTTGTTTCTCCACCCGAGGAAACCTTAAATGAACTGACGAAAGACAGAGA 300
DB 384 TCCTTTATTTGTTTCTCCACCCGAGGAAACCTTAAATGAACTGACGAAAGACAGAGA 443
QY 301 CAAATCCGTCAAGGCGTGAAGAACTCCAGAAAGTTCTGCGAGGAGAGACCTTACATG 360
DB 444 CAAATCCGTCAAGGCGTGAAGAACTCCAGAAAGTTCTGCGAGGAGAGACCTTACATG 503
QY 361 CATGAAGATTGAAAGGCGCAGTAGAGAGATTTATATGAAACAGACAAAGGATACAG 420
DB 504 CATGAAGATTGAAAGGCGCAGTAGAGAGATTTATATGAAACAGACAAAGGATACAG 563
QY 421 ACAGCCAGCGTCATCTGCTTTGACTGTAGTGAAGAACTCCATGAAGATCTCTTTTCTAT 480
DB 564 ACAGCCAGCGTCATCTGCTTTGACTGTAGTGAAGAACTCCATGAAGATCTCTTTTCTAT 623
QY 481 TCAGAGAGGAGGAGCTTAATAGGCTCGAGATCTTGGTGCAATTTTACGTGTTGGTGTG 540
DB 624 TCAGAGAGGAGGAGCTTAATAGGCTCGAGATCTTGGTGCAATTTTACGTGTTGGTGTG 683
QY 541 AAAAGATTTCAATGAGACACAGCTGCGCCGAGATTCGCGACAGTAAGATCATGTGTTCC 600
DB 684 AAAAGATTTCAATGAGACACAGCTGCGCCGAGATTCGCGACAGTAAGATCATGTGTTCC 743
QY 601 GTGAATGACGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAGTCTCTGC 660
DB 744 GTGAATGACGCTTTCAGGCTCTGCAAGGCATCATCTCAATTTTGAAGAGTCTCTGC 803
QY 720 ATCGAATTCATGACGTGAACCATTCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720
DB 804 ATCGAATTCATGACGTGAACCATTCACATATGTGAGAGAGTCAATTTCAAGTTGTC 863
QY 721 GTGAGAGGAAACGCGTTCCGACATGCGCGCAACGTGACAGGCTCTCTGCAGCTTCAAG 780
DB 864 GTGAGAGGAAACGCGTTCCGACATGCGCGCAACGTGACAGGCTCTCTGCAGCTTCAAG 923
QY 840 ATCAATGACTCGGTCACTCAATGAAAGCCCTTTCTGTGTGAAGACACTTAATTACTG 840
DB 924 ATCAATGACTCGGTCACTCAATGAAAGCCCTTTCTGTGTGAAGACTTAATTACTG 983
QY 900 TGTCCAGGCGCTATCTTAAAGAAAGTGGCATGAAGGTCAGCTCAGGTGAGTAAG 900
DB 944 TGTCCAGGCGCTATCTTAAAGAAAGTGGCATGAAGGTCAGCTCAGGTGAGTAAG 1043
QY 960 GATGCGCTCTCTTTATCTCAGTTCTGTATCATATCACCACCACTGTTCTGACGGT 960
DB 1044 GATGCGCTCTCTTTATCTCAGTTCTGTATCATATCACCACCACTGTTCTGACGGT 1103
QY 1020 TCCATCTGCGCATGCGCTGCTGATCTGTTCTGTGCTTACGCTGCTCTCTCTG 1020

```

```

DB 1104 TCCATCTGCGCATGCGCTGCTGATCTGTCTCTCTCCAGCCCTGCTCTCTCTG 1163
QY 1021 TGGTTCTGCGCCCTCTCTGCTCACTGTGATTATCAAGAGGATCTCTCCAGCCCTGCGAG 1080
DB 1164 TGGTTCTGCGCCCTCTCTGCTCACTGTGATTATCAAGAGGATCTCTCCAGCCCTGCGAG 1223
QY 1081 GAGAGTGAGGAAAA 1094
DB 1224 GAGAGTGAGGAAAA 1237

RESULT 4
ABV73882
ID ABV73882 standard; cDNA; 5540 BP.
AC ABV73882;
XX 08-JAN-2003 (first entry)
DT
XX Human anthrax toxin receptor cDNA.
DE Anthrax; toxin; receptor; human; TEM8; antibacterial; gene; ss.
KW Homo sapiens.
XX Location/Qualifiers
XX Key 144..1838
XX CDS 144..1838
FT sig_peptide 144..224
FT /*cag= b
FT mat_peptide 225..1838
FT /*cag= c
PN MO200246228-A2.
XX 13-JUN-2002.
PD
XX 03-OCT-2001; 2001WO-US030941.
XX 05-DEC-2000; 2000US-0251481P.
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX Young JAT, Bradley KA, Collier RJ, Mogridge JS;
XX WPI: 2002-713235/77.
XX P-PSDB; ABP54904.
XX Novel isolated polypeptide useful for identifying agent that prevents or
XX reduces effect of anthrax toxin on host cell, for treating human or non-
XX human animal suffering from anthrax.
XX Disclosure; Page 32-37; 45pp; English.
XX
XX The present sequence is that of cDNA encoding a human anthrax toxin
XX receptor (ATR). The cDNA, previously designated TEM8, was originally
XX isolated by PCR from HeLa cell and human placenta cDNA libraries. It was
XX identified in a database screening using a newly isolated human cDNA (see
XX ABV73881) which encodes an ATR (see ABP54903) that is identical to the
XX TEM8 polypeptide except in the cytoplasmic tail, suggesting differential
XX splicing of a primary mRNA transcript. TEM8 was not previously identified
XX as an ATR. The invention provides ATR polypeptides and polynucleotides,
XX host cells, vectors, and transgenic and knock-out animals. It also
XX provides methods for identifying molecules that bind the ATR and which
XX reduce the toxicity of anthrax toxin. A claimed method for treating
XX anthrax in a human or animal involves administering an agent that
XX inhibits binding between anthrax toxin protective antigen (PA) and ATR at
XX a level effective to reduce the severity of anthrax. Suitable agents
XX include the TEM8 polypeptide or a PA-binding fragment of it, a PA-binding
XX polypeptide at least 80% identical to these, a fusion protein, a
XX monoclonal or polyclonal antibody, a polysaccharide, a lipid or a nucleic
XX acid. ATR polynucleotides can also be used in the recombinant production

```



OY	1	ATGGCCAGCGGCGAGCGGAGAGCCCTCGGACATCGGCTTCAGTGGCTCTCTTTGGCACT	60
Db	144	ATGGCGACGGCGGAGCGGAGAGCCCTCGGACATCGGCTTCAGTGGCTCTCTTTGGCACT	203
OY	61	CTGGGCTCATCTGTGGCGCGGCGAAGGGGAGACGACAGAGAGATGGGGGTCCAGCTGGTAC	120
Db	204	CTGGGCTCATCTGTGGCGCGGCGAAGGGGAGACGACAGAGAGATGGGGGTCCAGCTGGTAC	263
OY	121	GGCGGATTTGACTGTACTTTCATTTTGGACAATCAGAAAGTGTCTGCACCATCTGAAT	180
Db	264	GGCGGATTTGACTGTACTTTCATTTTGGACAATCAGAAAGTGTGTCTGCACCATCTGAAT	323
OY	181	GAATCTATTACTTTGTGGACAAGTTGGCTCAAAATTCAACAGCCCACTGTGGAAATG	240
Db	324	GAATCTATTACTTTGTGGACAAGTTGGCTCAAAATTCATCAGCCCACTGTGGAAATG	383
OY	241	TCCTTTATTTGTTTCTCCACCCGAGAAACAACCTTAATGAACCTACAGAAACAGAGAA	300
Db	384	TCCTTTATTTGTTTCTCCACCCGAGAAACAACCTTAATGAACCTACAGAAACAGAGAA	443
OY	301	CAAAATCCGTCAAGGCGCTGAGAAAGACTCGAAGAAATTCTGCCAGAGAGACACTTACATG	360
Db	444	CAAAATCCGTCAAGGCGCTGAGAAAGACTCGAAGAAATTCTGCCAGAGAGACACTTACATG	503
OY	361	CATGAAGATTTGAAAAGGCCCGCTGAGCAGATTTATTGAAAACAGACAAAGGGTACAG	420
Db	504	CATGAAGATTTGAAAAGGCCCGCTGAGCAGATTTATTGAAAACAGACAAAGGGTACAG	563
OY	421	ACAGCAGGTCATCATATTGCTTGACTGATGGAGAACTCAATGAAGATCTCTTTTCTAT	480
Db	564	ACAGCAGGTCATCATATTGCTTGACTGATGGAGAACTCAATGAAGATCTCTTTTCTAT	623
OY	481	TCAGAGAGGAGGCTAATATGATCTCGAAGATCTTGTGTGCAATTTGTTACTGTGTGTGTG	540
Db	624	TCAGAGAGGAGGCTAATATGATCTCGAAGATCTTGTGTGCAATTTGTTACTGTGTGTGTG	683
OY	541	AAAGATTTTCATGAGACACAGCTGGGCCGCGAATGGCGACAGTAAAGATCATGTGTTCCC	600
Db	684	AAAGATTTTCATGAGACACAGCTGGGCCGCGAATGGCGACAGTAAAGATCATGTGTTCCC	743
OY	601	GTGAATGACGGCTTTCAGGCTCTGAGAGGCATCATCTCAATTTTGAAGAGTCTGCG	660
Db	744	GTGAATGACGGCTTTCAGGCTCTGAGAGGCATCATCTCAATTTTGAAGAGTCTGCG	803
OY	661	ATCGAAATTTCTAGCAGCTGAACCATTCACCATATGTGTGCGAGAGTCAATTCAGTTGTC	720
Db	804	ATCGAAATTTCTAGCAGCTGAACCATTCACCATATGTGTGCGAGAGTCAATTCAGTTGTC	863
OY	721	GTGAGAGGAAACGGCTTCCGACATGCCCGGCAACGTGACAGGGTCTCTGCAGCTTCAAG	780
Db	864	GTGAGAGGAAACGGCTTCCGACATGCCCGGCAACGTGACAGGGTCTCTGCAGCTTCAAG	923
OY	781	ATCAATGACTCGGTGACACTCAATGAGAAAGCCCTTTCGTGGAAGATCTTAATTACG	840
Db	924	ATCAATGACTCGGTGACACTCAATGAGAAAGCCCTTTCGTGGAAGATCTTAATTACG	983
OY	841	TGTCAGACGCTATCTTAAAGAAAGTTGGCATGAAGCTGCACTCGAGTCCAGATGAAAC	900
Db	984	TGTCAGACGCTATCTTAAAGAAAGTTGGCATGAAGCTGCACTCGAGTCCAGATGAAAC	1043
OY	901	GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCATCAACACACACACTGTTCTGACGGT	960
Db	1044	GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCATCAACACACACACTGTTCTGACGGT	1103
OY	961	TCCATCTCTGGCCATGGCCCTGCTGATCCGTTCCTGCTCTGCTCTGAGCCCTGCTCTCTGG	1020
Db	1104	TCCATCTCTGGCCATGGCCCTGCTGATCCGTTCCTGCTCTGCTCTGAGCCCTGCTCTCTGG	1163
OY	1021	TGGTCTGGCCCTCTGCTGACCTGTGATTAATCAAGAGAGTCCCTCCACCCCTGCGAG	1080
Db	1164	TGGTCTGGCCCTCTGCTGACCTGTGATTAATCAAGAGAGTCCCTCCACCCCTGCGAG	1222
OY	1081	GAGAGTGAAGAAAAA	1094

Db 1224 GAGAGTGAGGAAGA 1237

RESULT 6  
ABX72003  
ID ABX72003 standard; DNA; 5540 BP.

DT 12-MAR-2003 (first entry)

DE DNA encoding human tumour endothelial marker TEM 8.

Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;

KM pan-endothelial marker; polycystic kidney disease; psoriasis;

kw diabetic retinopathy; insulinoma alcoholism; cannot angiogenesis; neoangiogenesis; immune response; cytostatic; antidiabetic; gene;

ophtalmolog  
KW  
XX

OS Homo sapiens.  
XX

PN	WO200283874
YY	

PD 24-OCT-2002.  
YY

PF 10-APR-2002; 2002WO-US008253.

PR 11-APR-2001; 2001US-0282850P.

XX

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

[illegible]

DR P-PSDB; ABU54430.

PT New purified human transmembrane protein, designated as tumor endothelial

PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or

PROLIFABIS.  
P1  
XX

PS Disclosure; page 117-120; 3/4pp; English.  
XX  
XX

CC The present invention relates to a novel method for the isolation of  
CC endothelial cells (ECs), and the identification of genes expressed in

CC normal and tumour ECs. Tumour endothelial marker (TEM), normal CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are

CC identified in human ECs. The human EC marker proteins and the  
CC polypeptide sequences encoding them are useful for detecting.

CC diagnosing or treating tumours as well as polycystic kidney disease,  
CC diabetic retinopathy, rheumatoid arthritis and psoriasis. They are

CC useful for inhibiting neoangiogenesis or tumour angiogenesis, for  
CC inducing an immune response to tumour endothelial cells in a patient

CC For identifying candidate drugs for treating tumours. The present

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]

	Best Local Similarity	Pred. No. 0;
99.8%;		

MARLINER  
T-024,  
COMBETALVAC  
C,  
MILITARY  
F,  
F-060  
S/  
S/Sgt

[illegible]

DB 144 ATGGCCACGGCGGAGCGGAAGCCCTCGGCATCGCTCCAGTGGCTCTCTTGGCCATC 20

61 CTGGTCTCACTCTGCGCCGGCAAGGGGACCGAGGAGATGGGGTCCAGCCCTGCTAC 120

Db 204 CTGCTGCTCATCTGCGCCGGCAAGGGGACGCAGGGAGGATGGGGTCCAGCTGCTAC 205

```
Qy 121 GCGCGATTGACCTGTACTTCTCATTTTGGACAAATCAGAAAGTGTGCTGACCACCTGGAAT 180
Db 264 GCGCGATTGACCTGTACTTCTCATTTTGGACAAATCAGAAAGTGTGCTGACCACCTGGAAT 323
Qy 181 GAATCTATTCTTTGTGTGGAAAGTGTGCTGACAAATTCATCAGCCCAAGTTTGAGATG 240
Db 324 GAATCTATTCTTTGTGTGGAAAGTGTGCTGACAAATTCATCAGCCCAAGTTTGAGATG 383
Qy 241 TCTTTATTGTTTCTCCACCGAGGAACAACCTTAATGAAATCTGACAGAAACAGAGAA 300
Db 384 TCTTTATTGTTTCTCCACCGAGGAACAACCTTAATGAAATCTGACAGAAACAGAGAA 443
Qy 301 CAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTCTCCAGAGAGAGACATTACATG 360
Db 444 CAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTCTCCAGAGAGAGACATTACATG 503
Qy 361 CATGAAGATTGAAAGGCGCCAGTGAAGCATTTTATATGAAAAACGACAGGGTACAGG 420
Db 504 CATGAAGATTGAAAGGCGCCAGTGAAGCATTTTATATGAAAAACGACAGGGTACAGG 563
Qy 421 ACAGCCAGCGTCATCATTTGCTTTTGACTGATGAGAACTCCATGAAAGATCTCTTTTCTAT 480
Db 564 ACAGCCAGCGTCATCATTTGCTTTTGACTGATGAGAACTCCATGAAAGATCTCTTTTCTAT 623
Qy 481 TCAGAGAGGAGGCTTAATAGGCTTCGAGATCTTGTGTGCAATGTTTACTGTGTGTGTG 540
Db 624 TCAGAGAGGAGGCTTAATAGGCTTCGAGATCTTGTGTGCAATGTTTACTGTGTGTGTG 683
Qy 541 AAAATTTTCATGAGACACAGCTGCGCCGGAATGGCGACAGTAAGATCATGTGTTTCCC 600
Db 684 AAAATTTTCATGAGACACAGCTGCGCCGGAATGGCGACAGTAAGATCATGTGTTTCCC 743
Qy 601 GTGAATGACGCGCTTTCAGAGCTCTGTGCAAGGCATCATCTCAATTTTGAAGAAGCTTCC 660
Db 744 GTGAATGACGCGCTTTCAGAGCTCTGTGCAAGGCATCATCTCAATTTTGAAGAAGCTTCC 803
Qy 661 ATCGAAATTTAGAGAGCTGAACCATTCACCATATGTGAGAGAGAGTCAATTTCAAGTTGTC 720
Db 804 ATCGAAATTTAGAGAGCTGAACCATTCACCATATGTGAGAGAGAGTCAATTTCAAGTTGTC 863
Qy 721 GTGAGAGGAAACGCGCTTCCGACATGCCCCGCAAGTGGACAGGGTCTTGTGACGCTTCAAG 780
Db 864 GTGAGAGGAAACGCGCTTCCGACATGCCCCGCAAGTGGACAGGGTCTTGTGACGCTTCAAG 923
Qy 781 ATCAATGACGCTGTCACACTCAATGAGAAAGCCCTTTTGTGGAAGACACTTATTTACTG 840
Db 924 ATCAATGACGCTGTCACACTCAATGAGAAAGCCCTTTTGTGGAAGACTTATTTACTG 983
Qy 841 TGTCCAGCGCTTATCTTAAAGAAAGTGGCATGAAAGCTGCACTTCAGGTGAGCATGAAAC 900
Db 984 TGTCCAGCGCTTATCTTAAAGAAAGTGGCATGAAAGCTGCACTTCAGGTGAGCATGAAAC 1043
Qy 901 GATGGCGCTCTTTTATATCTCAAGTTTGTGTCATCATCACACCAACACACTGTTTGAAGGCT 960
Db 1044 GATGGCGCTCTTTTATATCTCAAGTTTGTGTCATCATCACACCAACACACTGTTTGAAGGCT 1103
Qy 961 TCCATCTGAGCATGCGCCCTGTGATCTGTTCCGCTCCTAGGCGGTGCTCTCTCTG 1020
Db 1104 TCCATCTGAGCATGCGCCCTGTGATCTGTTCCGCTCCTAGGCGGTGCTCTCTCTG 1163
Qy 1021 TGGTTCTGCGCCCTCTGTGTGACATGTGATTTATCAAGAGAGTCCCTCCACCCCTGCGAG 1080
Db 1164 TGGTTCTGCGCCCTCTGTGTGACATGTGATTTATCAAGAGAGTCCCTCCACCCCTGCGAG 1223
Qy 1081 GAGAGTGAAGAAA 1094
Db 1224 GAGAGTGAAGAA 1237
```

```
RESULT 7
ADR48215
ID ADR48215 standard; cDNA; 5540 BP.
XX
```

```
AC ADR48215;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human tumour endothelial marker 8 precursor encoding cDNA SEQ:3.
XX
KW pancreatic cancer-associated transcript; pancreatic cancer; human;
XX cytosolic; gene therapy; protein therapy;
XX tumour endothelial marker 8 precursor; TEM6; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 144..1838
FT /tag= a
FT /product= "tumour endothelial marker 8 precursor (TEM6)"
XX
XX MO2004074510-A1.
XX
XX 02-SEP-2004.
XX
XX 18-FEB-2004; 2004MO-AU000194.
XX
XX 18-FEB-2003; 2003AU-00900747.
XX
XX (GARV-) GARVAN INST MEDICAL RES.
XX
XX Binklin A, Segara D, Henshall S, Sutherland R;
XX WPI; 2004-635591/61.
XX DR P-PSDB; ADR48216.
XX
XX
XX Detecting pancreatic cancer-associated transcript in a biological sample,
XX useful for diagnosing or treating the disease, comprises contacting the
XX sample with a polynucleotide that selectively hybridizes to a specific
XX sequence.
XX
XX Claim 70; SEQ ID NO 3; 263bp; English.
XX
XX
XX The present invention describes a method for detecting a pancreatic
XX cancer-associated transcript in a biological sample. The method comprises
XX contacting the biological sample with a polynucleotide that selectively
XX hybridizes to a sequence at least 80% identical to a sequence as shown in
XX any one of Tables 3 to 25 in the specification or having the GenBank
XX Accession Number AF279145. Also described: (1) diagnosing pancreatic
XX cancer in a human or animal subject being tested, determining the
XX likelihood that a subject having a pancreatic cancer will survive, or
XX determining the suitability of a subject having a pancreatic cancer for
XX surgical resection therapy; (2) detecting a pancreatic cancer-associated
XX polypeptide in a biological sample; (3) determining the likelihood that a
XX subject having a pancreatic cancer will survive; and (4) monitoring the
XX efficacy of a therapeutic treatment of pancreatic cancer. A pancreatic
XX cancer-associated transcript has cytosolic activity, and can be used in
XX gene and protein therapy. A pancreatic cancer-associated transcript
XX polynucleotide, a vector comprising the polynucleotide, an isolated
XX polypeptide or an antibody that binds to the isolated polypeptide can be
XX used for diagnosing or prognosing pancreatic cancer or for preparing a
XX medicament for the treatment of pancreatic cancer. The prognostic or
XX diagnostic methods are useful for the early detection of pancreatic
XX cancer or its metastases, and for monitoring the progress of disease such
XX as during remission or following surgery or chemotherapy. The present
XX sequence encodes human tumour endothelial marker 8 precursor (TEM6),
XX which is used in the exemplification of the present invention.
XX
XX Sequence 5540 BP; 1601 A; 1306 C; 1187 G; 1446 T; 0 U; 0 Other;
XX
```

```
Query Match 98.8%; Score 1090.8; DB 13; Length 5540;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCCACGGCGAGAGCCGCTTGGCATGAGGCTTCCATGAGGCTCTCTTTGGCCACT 60
Db 144 ATGGCCACGGCGAGAGCCGCTTGGCATGAGGCTTCCATGAGGCTCTCTTTGGCCACT 203
```

QY 61 CTGGTCTCATCTGCGCGGCAAGGGGAGCGAGGGAGATGGGGGTCCAGCTTGCTAC 120  
 DB 204 CTGGTCTCATCTGCGCGGCAAGGGGAGCGAGGGAGATGGGGGTCCAGCTTGCTAC 263  
 QY 121 GGGCGATTTGACCTGTACTTCAATTTTGGACAAATCAGAAAGTGTGTCCACTCGAAT 180  
 DB 264 GGGCGATTTGACCTGTACTTCAATTTTGGACAAATCAGAAAGTGTGTCCACTCGAAT 323  
 QY 181 GAATCTATTACTTTGGGAAACAGTTGTCTCAAAATTCATCAGCCCACTGTAGAAATG 240  
 DB 324 GAATCTATTACTTTGGGAAACAGTTGTCTCAAAATTCATCAGCCCACTGTAGAAATG 383  
 QY 241 TCTTTATTTGTTTCTCCACCGGAGAACTTAATGAAGTGAAGTGAAGAGAGAGAA 300  
 DB 384 TCTTTATTTGTTTCTCCACCGGAGAACTTAATGAAGTGAAGTGAAGAGAGAGAA 443  
 QY 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTAATG 360  
 DB 444 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTAATG 503  
 QY 361 CATGAAGATTTGAAGGCGCAGTGAAGATTTATTAAGAAACAGCAAGGGTACAGG 420  
 DB 504 CATGAAGATTTGAAGGCGCAGTGAAGATTTATTAAGAAACAGCAAGGGTACAGG 563  
 QY 421 ACAGCCAGGCTCATCATTTGCTTGTGATGAGAGAACTCATGAAGATCTTTTTCAT 480  
 DB 564 ACAGCCAGGCTCATCATTTGCTTGTGATGAGAGAACTCATGAAGATCTTTTTCAT 623  
 QY 481 TCAAGAGAGAGGCTTAATAGTCTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 540  
 DB 624 TCAAGAGAGAGGCTTAATAGTCTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 683  
 QY 541 AAAGATTTCAATGAGACACAGTGGCGGAGTTGGGAGCAGTAAGGATCATGTGTTCC 600  
 DB 684 AAAGATTTCAATGAGACACAGTGGCGGAGTTGGGAGCAGTAAGGATCATGTGTTCC 743  
 QY 601 GTGATGACGGCTTTCAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTTGC 660  
 DB 744 GTGATGACGGCTTTCAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTTGC 803  
 QY 661 ATCGAATTTCAAGAGCTGAACCATCCACATATGTGAGAGAGAGTCAATTCAGTTGTC 720  
 DB 804 ATCGAATTTCAAGAGCTGAACCATCCACATATGTGAGAGAGAGTCAATTCAGTTGTC 863  
 QY 721 GTGAGAGGAAAGGCTTCCGACATGCGCGCAACGAGCAGGGTCTGTGAGGCTTCAAG 780  
 DB 864 GTGAGAGGAAAGGCTTCCGACATGCGCGCAACGAGCAGGGTCTGTGAGGCTTCAAG 923  
 QY 781 ATCAATGACCTGCTCACACTCAATGAGAAAGCCCTTTCTGTGAAGACACTTATTTACTG 840  
 DB 924 ATCAATGACCTGCTCACACTCAATGAGAAAGCCCTTTCTGTGAAGACACTTATTTACTG 983  
 QY 841 TGTCCAGGCGCTATTTTAAAGAGAGTGGCATGAAGCTGCACTCCAGTCAAGCATGAAC 900  
 DB 984 TGTCCAGGCGCTATTTTAAAGAGAGTGGCATGAAGCTGCACTCCAGTCAAGCATGAAC 1043  
 QY 901 GATGGCCCTCTCTTTATCTCAGTCTGTCAATCATCACACACACACTGTTTGAAGGCT 960  
 DB 1044 GATGGCCCTCTCTTTATCTCAGTCTGTCAATCATCACACACACACTGTTTGAAGGCT 1103  
 QY 961 TCCATCTTGCCCATGCGCCCTGTGATCTGTTCTGCTCTAGCCCTGCTCTCTCTG 1020  
 DB 1104 TCCATCTTGCCCATGCGCCCTGTGATCTGTTCTGCTCTAGCCCTGCTCTCTCTG 1163  
 QY 1021 TGGTTTGGCCCTCTGTGCTGCACTGTATTAAGAGAGTCTCTTCCACCCCTGCGAG 1080  
 DB 1164 TGGTTTGGCCCTCTGTGCTGCACTGTATTAAGAGAGTCTCTTCCACCCCTGCGAG 1223  
 QY 1081 GAGAGTGAGAGAAA 1094  
 DB 1224 GAGAGTGAGAGAAA 1237

RESULT 8  
 AAD05303  
 ID AAD05303 standard; cDNA; 2447 BP.  
 XX  
 AC AAD05303;  
 XX  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE Human secreted protein-encoding gene 4 cDNA clone HMLFR02, SEQ ID NO:14.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification;  
 KW chromosome 19; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 140..1351  
 FT /\*tag= a  
 FT /product= "Human secreted protein"  
 FT sig\_peptide 140..220  
 FT /\*tag= b  
 FT mat\_peptide 221..1348  
 FT /\*tag= c  
 FT /product= "Mature human secreted protein"  
 FT  
 XX W0200134626-A1.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PE 01-NOV-2000; 2000MO-US030045.  
 XX  
 PR 05-NOV-1999; 99US-0163581P.  
 PR 30-JUN-2000; 2000US-0215133P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;  
 XX WPI: 2001-308778/32.  
 DR P-PSDB; AAE01439.  
 DR  
 XX  
 PT New nucleic acid molecules encoding 28 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 PT  
 PS Claim 1; Page 425-426; 562pp; English.  
 XX  
 CC AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
 CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
 CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating,  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 28 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,











PS Example; SEQ ID NO 17; 64pp; English.  
 XX  
 XX The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the vWF. Also described are: a method of preventing or  
 CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to *Bacillus anthracis*;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This sequence encodes a  
 CC fusion protein comprising mouse TNFGS197 and immunoglobulin (Ig) that can  
 CC be used to treat exposure to or prevent a symptom of anthrax.  
 XX  
 XQ Sequence 1674 BP; 433 A; 461 C; 424 G; 356 T; 0 U; 0 Other;

Query Match	87.0%	Score 960.8	DB 11	Length 1674
Best Local Similarity	99.8%	Pred. No. 3.9e-284		
Matches 962; Conservative	0	Mismatches 2	Indels 0	Gaps 0

OY	1	ATGGCACAAGGCGGAGCGGAGAGAGCCCTCCGGCATTCGCGCTTCCAGAGTGGCTCTCTTTGGGCACT	60
Db	13	ATGGCACAAGGCGGAGCGGAGAGAGCCCTCCGGCATTCGCGCTTCCAGAGTGGCTCTCTTTGGGCACT	72
OY	61	CTGATGCTCATTCCTGCGCCGGGCAAGGAGGAGACGACGGAGAGATGGGGGTCCAGCTTCAC	120
Db	73	CTGATGCTCATTCGCGCCGGGCAAGGAGGAGACGACGGAGAGATGGGGGTCCAGCTTCAC	132
OY	121	GGCGGAATTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTGCTGCACCACTGGAAAT	180
Db	133	GGCGGAATTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTGCTGCACCACTGGAAAT	192
OY	181	GAAATCTATTAATCTTTGGGAACAGTTGGCTCAACAAATTCATCGACCCCAAGTTGAAATG	240
Db	193	GAAATCTATTAATCTTTGGGAACAGTTGGCTCAACAAATTCATCGACCCCAAGTTGAAATG	252
OY	241	TCCTTTATTTGTTTCTCCACCCGAGGACAACCTTAATGAAACTGACAGAGACAGAGAA	300
Db	253	TCCTTTATTTGTTTCTCCACCCGAGGACAACCTTAATGAAACTGACAGAGACAGAGAA	312
OY	301	CAAAATCCGTCAGGCTTAAGAAATCTCCAGAAAGTTCTGCCAGAGAGAGACAATTTACATG	360
Db	313	CAAAATCCGTCAGGCTTAAGAAATCTCCAGAAAGTTCTGCCAGAGAGAGACAATTTACATG	372
OY	361	CATGAAGGATTTGAAAGGGCCAGTGGAGCAGATTATATGAAACACAGCAAGGGTACAGG	420
Db	373	CATGAAGGATTTGAAAGGGCCAGTGGAGCAGATTATATGAAACACAGCAAGGGTACAGG	432
OY	421	ACAGCCAGGCTCATCATCTTCTTTGACTGATGAGAACTCCATGAAAGTCTCTTTTCTAT	480
Db	433	ACAGCCAGGCTCATCATCTTCTTTGACTGATGAGAACTCCATGAAAGTCTCTTTTCTAT	492
OY	481	TCAGAGAGGAGAGGCTTAATAGTCTCGAGATCTTGGTGCAATGCTTAACTGTGTGGGTG	540
Db	493	TCAGAGAGGAGAGGCTTAATAGTCTCGAGATCTTGGTGCAATGCTTAACTGTGTGGGTG	552
OY	541	AAAGATTTCAATGAGACACAGCTGGGCCCGGATTGCGGACGTAAAGATCATGTGTTCCC	600
Db	553	AAAGATTTCAATGAGACACAGCTGGGCCCGGATTGCGGACGTAAAGATCATGTGTTCCC	612
OY	601	GTTAATGACGCGCTTCAGGCTCTGCAAGGCATCATCOACTCAATTTTGAAGAGTCTGTC	660
Db	613	GTTAATGACGCGCTTCAGGCTCTGCAAGGCATCATCOACTCAATTTTGAAGAGTCTGTC	672
OY	661	ATCGAAATCTAGAGACTGAACCATCACCATATGTCAGGAGAGATCTTCAAGTGTGC	720
Db	673	ATCGAAATCTAGAGACTGAACCATCACCATATGTCAGGAGAGATCTTCAAGTGTGC	732
OY	721	GTTAGAGGAAACGGCTTCGACACATGCCCAGCAAGTGGACAGGGTCTCTGCACTTCAAG	780
Db	733	GTTAGAGGAAACGGCTTCGACACATGCCCAGCAAGTGGACAGGGTCTCTGCACTTCAAG	792
OY	781	ATCAATGACTCGGTACACTCAATGAAAGCCCTTTTCTGTGGAAGACACTTAATTTACTG	840

	793	841	853	901	913	961	973
Db	ATCATGACTCGTCACTCAATGAGAAGCCCTTTCTGTGGAAGATACTATTACTG						
Qy	TGTCAGGGCTTACTTTAAAGAAGTTGGCATGAAAGCTGCACCTCAGGTGCATGAAAC						
Db	TGTCAGGGCCATCTTTAAAGAAAGTTGGCATGAAAGCTGCACCTCAGGTGCATGAAAC						
Qy	GATGGCCCTCTCTTTATCTCCAGTTCTGTCAATCATCACCAACAACACTGTTCTGACGGT						
Db	GATGGCCCTCTCTTTATCTCCAGTTCTGTCAATCATCACCAACAACAACACTGTTCTGACGGT						
Qy	TCCA	964					
Db	CCCA	976					

RESULT 12  
AD100545  
ID AD100545 standard, DNA, 1650 BP.  
XX  
AC AD100545;  
XX  
DT 22-APR-2004 (first entry)

DE	Human TANGO 197 Ig FcR mutated fusion DNA - plasmid pKTOX127.
XX	fusion; von Willebrand factor A-like domain; vWF; antibacterial;
XX	cutaneous; inhalation anthrax; human; TANGO 197 Ig FcR fusion; mutant;
KM	Plasmid pKTOX127; ds; gene.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PN	US2003144193-A1.
XX	
PD	31-JUL-2003.
XX	
PF	24-JUL-2002; 2002US-00201292.
XX	
PR	20-DEC-2001; 2001US-00038307.
XX	
PA	(ROTT/) ROTTMAN J B.
PA	(OKEE/) O'KEEFE T L.
PA	(OZKA/) OZKAYNAK E.
PA	(HEAL/) HEALEY J J.
PI	Rotman JB, O'Keefe TL, Ozkaynak E, Healey JJ;
XX	
DR	WPI; 2003-720708/68.
XX	P-PSDB; ADI00546.
PT	New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or
PT	ameliorating symptoms of anthrax comprises a von Willebrand factor A-like
PT	domain (vWF) amino acid sequence and an amino acid sequence heterologous
PT	to the vWF.
PS	
XX	Example; SEQ ID NO 13; 86pp; English.
XX	
CC	The invention relates to a novel fusion polypeptide comprising a von
CC	Willebrand factor A-like domain (vWF) amino acid sequence and an amino
CC	acid sequence heterologous to the vWF. The polypeptide of the invention
CC	demonstrates antibacterial activities whilst the composition and method
CC	may be useful in preventing or ameliorating the symptoms of cutaneous
CC	and/or inhalation anthrax. The current sequence is that of the human
CC	TANGO 197 Ig FcR mutated fusion DNA of the invention.
XX	
SQ	Sequence 1650 BP; 427 A; 454 C; 419 G; 350 T; 0 U; 0 Other;
Query Match	86.3%; Score 952.4; DB 10; Length 1650;
Best Local Similarity	99.4%; Pred.No. 1.5e-281;
Matches 956; Conservative	0; Mismatches 6; Indels 0; Gaps 0;
1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCAATCGGCTTCCAGTGGCTCTCTTGGCCACT	60

```

Db 1 ATGGCCAGGCGGAGGAGAGCCCTTCGACATCGGCTTCAGTGGCTCTCTTTGGCCACT 60
Qy 61 CTGGTGCTCATCTGCGCGGCGGAGGAGGAGGAGATGAGGAGTCCAGCTGCTAC 120
Db 61 CTGGTGCTCATCTGCGCGGCGGAGGAGGAGGAGATGAGGAGTCCAGCTGCTAC 120
Qy 121 GCGGAGTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180
Db 121 GCGGAGTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCCAAGTTGGAATG 240
Db 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCCAAGTTGGAATG 240
Qy 241 TCTTTATTTGTTTCTCCACCGAGGAACAACCTTAATGAAACATGACAGGAAGAGAGAA 300
Db 241 TCTTTATTTGTTTCTCCACCGAGGAACAACCTTAATGAAACATGACAGGAAGAGAGAA 300
Qy 301 CAAATCCGCTCAAGGCTGAGAAAGCTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 360
Db 301 CAAATCCGCTCAAGGCTGAGAAAGCTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 360
Qy 361 CATGAAAGATTTGAAAGGCGGAGGAGGAGGAGATTTATTTGAAACAGCAAGGGTACAG 420
Db 361 CATGAAAGATTTGAAAGGCGGAGGAGGAGGAGATTTATTTGAAACAGCAAGGGTACAG 420
Qy 421 ACAGCAGGCTCATCTGTTTGAATGATGAGAGAACTCCAGAAAGCTCTTTTTCAT 480
Db 421 ACAGCAGGCTCATCTGTTTGAATGATGAGAGAACTCCAGAAAGCTCTTTTTCAT 480
Qy 481 TCAGAGAGGAGGCTTAATAGTCTCGAGATCTTGATGCAATTTGTTTACTGTGTGTG 540
Db 481 TCAGAGAGGAGGCTTAATAGTCTCGAGATCTTGATGCAATTTGTTTACTGTGTGTG 540
Qy 541 AAAATTTTCAATGAGACACAGCTGCGCGGATTTGGGAGAGATGATGTTTCCC 600
Db 541 AAAATTTTCAATGAGACACAGCTGCGCGGATTTGGGAGAGATGATGTTTCCC 600
Qy 601 GTGAATGACGCTTTCAGGCTCTGCAAGGAGCATCACTCAATTTTGAAGAAGCTCCG 660
Db 601 GTGAATGACGCTTTCAGGCTCTGCAAGGAGCATCACTCAATTTTGAAGAAGCTCCG 660
Qy 661 ATCGAAATTTCTAGAGAGTGAACATCCACATATGTGAGAGAGATCAATTTCAAGTGTG 720
Db 661 ATCGAAATTTCTAGAGAGTGAACATCCACATATGTGAGAGAGATCAATTTCAAGTGTG 720
Qy 721 GTGAGAGAAACGCTTCCGACATGCGCGCAACGTGACAGGCTCTCTGCAAGCTTCAAG 780
Db 721 GTGAGAGAAACGCTTCCGACATGCGCGCAACGTGACAGGCTCTCTGCAAGCTTCAAG 780
Qy 781 ATCAATGACTCGGTCACATCAATGAGAAAGCCCTTTCTGTGGAAGACATTAATTTACTG 840
Db 781 ATCAATGACTCGGTCACATCAATGAGAAAGCCCTTTCTGTGGAAGACATTAATTTACTG 840
Qy 841 TGTCACAGGCTTATCTTAAGAAAGTGTGCAATGAAAGTGTGCACTCCAGGTGACATGAAC 900
Db 841 TGTCACAGGCTTATCTTAAGAAAGTGTGCAATGAAAGTGTGCACTCCAGGTGACATGAAC 900
Qy 901 GATGAGCTCTTTTATCTCAAGTTCTGTCAATCAACACACACTGTTTCCGAA 960
Db 901 GATGAGCTCTTTTATCTCAAGTTCTGTCAATCAACACACACTGTTTCCGAA 960
Qy 961 TC 962
Db 961 TC 962

```

RESULT 13  
AD100547  
ID AD100547 standard; DNA; 1650 BP.  
XX AC AD100547;

```

XX 22-Apr-2004 (first entry)
DT Human TANGO 197 Ig FcR WT fusion DNA - plasmid pLTKR129.
XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;
DE cutaneous; inhalation anthrax; human; TANGO 197 Ig FcR fusion; wild-type;
XX plasmid pLTKR129; ds; gene.
XX Homo sapiens.
XX US2003144193-A1.
XX 31-Jul-2003.
XX 24-Jul-2002; 2002US-00201292.
XX 20-Dec-2001; 2001US-00038307.
XX (ROTT/) ROTTMAN J B.
XX (OKEE/) O'KEEFE T L.
XX (OZKA/) OZKAYNAK E.
XX (HEAL/) HEALEY J J.
XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;
XX WPI; 2003-720708/68.
XX P-PSDB; AD100548.
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or
XX ameliorating symptoms of anthrax comprises a von Willebrand factor A-like
XX domain (vWF) amino acid sequence and an amino acid sequence heterologous
XX to the vWF.
XX Example; SEQ ID NO 15; 86pp; English.
XX The invention relates to a novel fusion polypeptide comprising a von
XX Willebrand factor A-like domain (vWF) amino acid sequence and an amino
XX acid sequence heterologous to the vWF. The polypeptide of the invention
XX demonstrates antibacterial activities whilst the composition and method
XX may be useful in preventing or ameliorating the symptoms of cutaneous
XX and/or inhalation anthrax. The current sequence is that of the human
XX TANGO 197 Ig FcR wild-type fusion DNA of the invention.
XX Sequence 1650 BP; 427 A; 453 C; 419 G; 351 T; 0 U; 0 Other;
S0

```

Query Match 86.3%; Score 952.4; DB 10; Length 1650;  
Best Local Similarity 99.4%; Pred. No. 1.5e-281;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy 1 ATGGCCAGGCGGAGGAGGAGCCCTTCGACATCGGCTTCAGTGGCTCTCTTTGGCCACT 60
Db 1 ATGGCCAGGCGGAGGAGGAGCCCTTCGACATCGGCTTCAGTGGCTCTCTTTGGCCACT 60
Qy 61 CTGGTGCTCATCTGCGCGGCGGAGGAGGAGGAGATGAGGAGTCCAGCTGCTAC 120
Db 61 CTGGTGCTCATCTGCGCGGCGGAGGAGGAGGAGATGAGGAGTCCAGCTGCTAC 120
Qy 121 GCGGAGTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180
Db 121 GCGGAGTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCCAAGTTGGAATG 240
Db 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCCAAGTTGGAATG 240
Qy 241 TCTTTATTTGTTTCTCCACCGAGGAACAACCTTAATGAAACATGACAGGAAGAGAGAA 300
Db 241 TCTTTATTTGTTTCTCCACCGAGGAACAACCTTAATGAAACATGACAGGAAGAGAGAA 300
Qy 301 CAAATCCGCTCAAGGCTGAGAAAGCTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 360
Db 301 CAAATCCGCTCAAGGCTGAGAAAGCTCCAGAAAGTTCTGCGAGAGAGACATTAACATG 360

```

QY 361 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGACAGGTTACAGG 420  
Db 361 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGACAGGTTACAGG 420  
QY 421 ACAGCCAGCGCTCATCTGTTTGTGATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
Db 421 ACAGCCAGCGCTCATCTGTTTGTGATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
QY 481 TCAGAGAGGGAGGCTTAATAGGCTTCGAGATCTGTATGATGATGATGATGATGATG 540  
Db 481 TCAGAGAGGGAGGCTTAATAGGCTTCGAGATCTGTATGATGATGATGATGATGATG 540  
QY 541 AAAGATTTTCAATGAGACACAGCTGCGCCGAGTTGCGAGATGAGTAAAGATGATGATG 600  
Db 541 AAAGATTTTCAATGAGACACAGCTGCGCCGAGTTGCGAGATGAGTAAAGATGATGATG 600  
QY 601 GTGAATGACGGCTTTCAAGGCTTCTGCAAGGATCTATCCATCAATTTTGAAGAACTCTGC 660  
Db 601 GTGAATGACGGCTTTCAAGGCTTCTGCAAGGATCTATCCATCAATTTTGAAGAACTCTGC 660  
QY 661 ATCGAAATTTAGAGAGCTGAACCATCCACCATATGTCAGAGAGTCAATTTCAAGTTGTC 720  
Db 661 ATCGAAATTTAGAGAGCTGAACCATCCACCATATGTCAGAGAGTCAATTTCAAGTTGTC 720  
QY 721 GTGAGAGGAAACGGCTTCCGACATGCGCCGCAACGTGACAGGCTCTCTGACGCTTCAAG 780  
Db 721 GTGAGAGGAAACGGCTTCCGACATGCGCCGCAACGTGACAGGCTCTCTGACGCTTCAAG 780  
QY 781 ATCAATGATCTGTCACACTCAATGAGAAAGCTTTTCTGTGAAAGACATTAATTTACTG 840  
Db 781 ATCAATGATCTGTCACACTCAATGAGAAAGCTTTTCTGTGAAAGACATTAATTTACTG 840  
QY 841 TGTCCAGGCGCTATCTTAAGAAAGTTGGCATGAAGCTGACCTCCAGTCAGCATGAC 900  
Db 841 TGTCCAGGCGCTATCTTAAGAAAGTTGGCATGAAGCTGACCTCCAGTCAGCATGAC 900  
QY 901 GATGCGCTCTTTTATCTCCAGTTCTGTCAATCAACACACACTGTTTCTCCAAA 960  
Db 901 GATGCGCTCTTTTATCTCCAGTTCTGTCAATCAACACACACTGTTTCTCCAAA 960  
QY 961 TC 962  
Db 961 TC 962

RESULT 14  
ADM64589  
ID ADM64589 standard; DNA; 1650 BP.  
XX  
XX ADM64589;  
AC  
XX 03-JUN-2004 (first entry)  
XX  
XX  
XX TANGO197 extracellular domain-mutant IgG Fc fusion protein DNA #2.  
DE  
XX antibacterial; gene therapy;  
XX von Willebrand factor A-like domain amino acid sequence;  
XX WVF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
XX inhalation anthrax; human; TANGO197; Immunoglobulin G; IgG;  
XX fragment of crystallisation; Fc; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX Synthetic.  
XX  
XX US2003134786-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX

PA (ROT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-829643/77.  
DR P-PSDB; ADM64590.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (WVF) amino acid sequence and an amino acid sequence heterologous  
PT to the WVF.  
XX  
XX Example; SEQ ID NO 13; 644p; English.  
XX  
XX The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (WVF) amino acid sequence and an amino acid sequence  
CC heterologous to the WVF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to Bacillus anthracis;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This sequence encodes a  
CC fusion protein comprising human TANGO197 extracellular region, minus the  
CC C<sub>2</sub> residues closest to the transmembrane region, and immunoglobulin G  
CC (19g) fragment of crystallisation (Fc) with mutations I235A and G237A.  
XX  
SQ Sequence 1650 BP; 427 A; 454 C; 419 G; 350 T; 0 U; 0 Other;  
Query Match 86.3%; Score 952.4; DB 11; Length 1650;  
Best Local Similarity 99.4%; Pred. No. 1.5e-281;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 1 ATGGCCACGGCGGAGGAGAGGCGCTCGGATGGCTTCCAGTGGCTCTGTTGGCCACT 60  
Db 1 ATGGCCACGGCGGAGGAGGAGGCGCTCGGATGGCTTCCAGTGGCTCTGTTGGCCACT 60  
QY 61 CTGGTGCTCATCTGCGCGCGGCAAGGGGAGCGAGGAGATGGGGTCCAGCTGTCTAC 120  
Db 61 CTGGTGCTCATCTGCGCGCGGCAAGGGGAGCGAGGAGATGGGGTCCAGCTGTCTAC 120  
QY 121 GCGGATTTGACCTGTACTTCAATTTTGGACAAATCAGAAAGTGTCTGACCACTGGAAT 180  
Db 121 GCGGATTTGACCTGTACTTCAATTTTGGACAAATCAGAAAGTGTCTGACCACTGGAAT 180  
QY 181 GAAATCTAATTAATCTTGGGAAAGTTGGCTCACAAATTCATCGCCACAGTTGAGATG 240  
Db 181 GAAATCTAATTAATCTTGGGAAAGTTGGCTCACAAATTCATCGCCACAGTTGAGATG 240  
QY 241 TCCCTTAATGTTTCTTCCACCCGAGGAAACCTTAATGAAATGACAGAAACAGAGAA 300  
Db 241 TCCCTTAATGTTTCTTCCACCCGAGGAAACCTTAATGAAATGACAGAAACAGAGAA 300  
QY 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACCTTAATG 360  
Db 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACCTTAATG 360  
QY 361 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACCTTAATG 360  
Db 361 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACCTTAATG 360  
QY 361 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGACAGGTTACAGG 420  
Db 361 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGACAGGTTACAGG 420  
QY 421 ACAGCCAGCGCTCATCTGTTTGTGATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
Db 421 ACAGCCAGCGCTCATCTGTTTGTGATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
QY 481 TCAGAGAGGGAGGCTTAATAGGCTTCGAGATCTGTATGATGATGATGATGATGATG 540  
Db 481 TCAGAGAGGGAGGCTTAATAGGCTTCGAGATCTGTATGATGATGATGATGATGATG 540  
QY 541 AAAGATTTTCAATGAGACACAGCTGCGCCGAGTTGCGAGATGAGTAAAGATGATGATG 600  
Db 541 AAAGATTTTCAATGAGACACAGCTGCGCCGAGTTGCGAGATGAGTAAAGATGATGATG 600

D 541 AAAATTTCATGACACAGCTGCCCCGGAATTGCGACAGTAAGATCATGTGTTCCC 600  
Q 601 GTGAATGACGGCTTTCAGGCTCTGCAAGCATCATCTCAATTTTGAAGAAGTCTGC 660  
D 601 GTGAATGACGGCTTTCAGGCTCTGCAAGCATCATCTCAATTTTGAAGAAGTCTGC 660  
Q 661 ATCGAATTTCTAGACAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
D 661 ATCGAATTTCTAGACAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
Q 721 GTGAGAGGAAACGGCTTCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAG 780  
D 721 GTGAGAGGAAACGGCTTCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAG 780  
Q 781 ATCAATGACTCGTCACTCACTCAATGAGAGCCCTTTTCTGTGAAGACACTTAATTAATG 840  
D 781 ATCAATGACTCGTCACTCACTCAATGAGAGCCCTTTTCTGTGAAGACTTAATTAATG 840  
Q 841 TGTCACAGCGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTTCAGTCAGCATGAA 900  
D 841 TGTCACAGCGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTTCAGTCAGCATGAA 900  
Q 901 GATGCGCTCTTTTATCTGATCTGTCATCATCAACACACACACTGTTCTCCAAA 960  
D 901 GATGCGCTCTTTTATCTGATCTGTCATCATCAACACACACACTGTTCTCCAAA 960  
Q 961 TC 962  
D 961 TC 962

## RESULT 15

ADM64591 ID ADM64591 standard; DNA; 1650 BP.

ADM64591;

03-JUN-2004 (first entry)

TANGO197 extracellular domain-wild type IgG Fc fusion protein DNA.

XX antibody; gene therapy;  
XX von Willebrand factor A-like domain amino acid sequence;  
XX vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
XX inhalation anthrax; human; TANGO197; immunoglobulin G; IgG;  
XX fragment of crystallisation; Fc; gene; ds.

OS Homo sapiens.  
OS Synthetic.

US2003134786-A1.

17-JUL-2003.

20-DEC-2001; 2001US-00038307.

20-DEC-2001; 2001US-00038307.

PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.

Roctman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

WPI; 2003-829643/77.

P-PSDB; ADM64592.

PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.

PS Example; SEQ ID NO 15; 64pp; English.

XX The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to Bacillus anthracis;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This sequence encodes a  
CC fusion protein comprising human TANGO197 extracellular region, minus the  
CC DG residues closest to the transmembrane region, and immunoglobulin G  
CC (IgG) fragment of crystallisation (Fc).

SQ Sequence 1650 BP; 427 A; 453 C; 419 G; 351 T; 0 U; 0 Other;

Query Match 86.3%; Score 952.4; DB 11; Length 1650;

Best Local Similarity 99.4%; Pred. No. 1.5e-281;

Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Q 1 ATGSCACGCGGAGGAGAGCCCTGCGCATCGGCTTCCAGTGGCTCTTTGGCCACT 60  
D 1 ATGSCACGCGGAGGAGAGCCCTGCGCATCGGCTTCCAGTGGCTCTTTGGCCACT 60  
Q 61 CTGGTCTCATCTGCGCCGCGCAAGGCGGACGAGGAGATGCGGCTTCAGCTCTTAC 120  
D 61 CTGGTCTCATCTGCGCCGCGCAAGGCGGACGAGGAGATGCGGCTTCAGCTCTTAC 120  
Q 121 GCGGATTTGACCTGTACTTCTTATTTTGAACAAATCAGAAAGTGTGTGACCACTGGAAT 180  
D 121 GCGGATTTGACCTGTACTTCTTATTTTGAACAAATCAGAAAGTGTGTGACCACTGGAAT 180  
Q 181 GAAATCTTACTTGTGAGAACAGTTGGCTCAACAAATTCATCAGCCACAGTTGAGATG 240  
D 181 GAAATCTTACTTGTGAGAACAGTTGGCTCAACAAATTCATCAGCCACAGTTGAGATG 240  
Q 241 TCCTTTATTTGTTTCTCACCCGAGAACACTTAATGAAGTGCAGAGAGACACTTAACATG 300  
D 241 TCCTTTATTTGTTTCTCACCCGAGAACACTTAATGAAGTGCAGAGAGACACTTAACATG 300  
Q 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTAACATG 360  
D 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTAACATG 360  
Q 361 CATGAGATTTGAAAGGCGCAGTGCAGATTTATTTATGAAAACAGACAAAGGTTACAGG 420  
D 361 CATGAGATTTGAAAGGCGCAGTGCAGATTTATTTATGAAAACAGACAAAGGTTACAGG 420  
Q 421 ACAGCCAGCGTCATCTTGTGACTGATGAGAACTCCATGAAAGTCTTTTCTAT 480  
D 421 ACAGCCAGCGTCATCTTGTGACTGATGAGAACTCCATGAAAGTCTTTTCTAT 480  
Q 481 TCAGAGAGGAGGCTAATAGTCTCGAGATCTTGTCGAATTTGTTTCTGTGTG 540  
D 481 TCAGAGAGGAGGCTAATAGTCTCGAGATCTTGTCGAATTTGTTTCTGTGTG 540  
Q 541 AAAGATTTCAATGAGACACAGCTGGCCGGAATGCGGACGTAAGATCATGTGTTCCC 600  
D 541 AAAGATTTCAATGAGACACAGCTGGCCGGAATGCGGACGTAAGATCATGTGTTCCC 600  
Q 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCCACTCAATTTTGAAGAAGTCTGTC 660  
D 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCCACTCAATTTTGAAGAAGTCTGTC 660  
Q 661 ATCGAATTTCTAGACAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
D 661 ATCGAATTTCTAGACAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
Q 721 GTGAGAGGAAACGGCTTCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAG 780  
D 721 GTGAGAGGAAACGGCTTCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAG 780  
Q 781 ATCAATGACTCGTCACTCACTCAATGAGAGCCCTTTTCTGTGAGAGACACTTAATTAATG 840

Db 781 ATCAATGACTCGGTCACTCAATGAGAGCCCTTTCTGTGAAATACTTATTACTG 840  
QY 841 TGTCCAGCGCCTATCTTAAAGAGTTGGCATGAAAGCTGCACTCCAGTCAGCATGAAC 900  
Db 841 TGTCCAGCGCCTATCTTAAAGAGTTGGCATGAAAGCTGCACTCCAGTCAGCATGAAC 900  
QY 901 GATGGCCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACACTGTTCTGACGGT 960  
Db 901 GATGGCCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACACTGTTCTCCCAA 960  
QY 961 TC 962  
Db 961 TC 962

Search completed: June 14, 2005, 14:06:24  
Job time : 704 secs

Db 309 ATGCCACGGCGAGAGCCCTCGGCATCGGCTTCCAGTGCTCTTTGGCACT 368

```
QY 61 CTGGTCTCATCTGGCCGCGGGAAGGGAAGGAGATGGGGTTCACCTGCTAC 120
Db 369 CTGGTCTCATCTGGCCGCGGGAAGGGAAGGAGATGGGGTTCACCTGCTAC 428
QY 121 GCGGATTTGACCTGTAATCTTATTTTGACAATCAGAAAGTGTGTGACCACTGGAAT 180
Db 429 GCGGATTTGACCTGTAATCTTATTTTGACAATCAGAAAGTGTGTGACCACTGGAAT 488
QY 181 GAAATCTATTAATCTTTGTGGAACAAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGAAATG 240
Db 489 GAAATCTATTAATCTTTGTGGAACAAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGAAATG 548
QY 241 TCCCTTATTTGTTTCTCCACCGGGAACAACCTTAATGAATCTGACGAAACAGAA 300
Db 549 TCCCTTATTTGTTTCTCCACCGGGAACAACCTTAATGAATCTGACGAAACAGAA 608
QY 301 CAAATCCGTCAGAGCCTAGAAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 360
Db 609 CAAATCCGTCAGAGCCTAGAAAGAACTCCAGAAAGTTCTGCGAGAGAGACACTTACATG 668
QY 361 CATGAAGATTTGAAAAGGCGCAGTGAAGATTTATTTATGAAAACAGAACAGGTTACAG 420
Db 669 CATGAAGATTTGAAAAGGCGCAGTGAAGATTTATTTATGAAAACAGAACAGGTTACAG 728
QY 421 ACAGCCAGGTCATGATTTGCTTTGACTGATGGAAGCTCCAGAAAGATCTTTTCTAT 480
Db 729 ACAGCCAGGTCATGATTTGCTTTGACTGATGGAAGCTCCAGAAAGATCTTTTCTAT 788
QY 481 TCAGAGAGGAGGCTAATAGATCTGAGATCTTGGTGCATTTGTTTACCTGTTGGTGTG 540
Db 789 TCAGAGAGGAGGCTAATAGATCTGAGATCTTGGTGCATTTGTTTACCTGTTGGTGTG 848
QY 541 AAAGATTTCAATGAGACACAGCTGGCCGGAATGGGACAGTAAGATCATGTGTTTCCC 600
Db 849 AAAGATTTCAATGAGACACAGCTGGCCGGAATGGGACAGTAAGATCATGTGTTTCCC 908
QY 601 GTGAATGACGGCTTCAAGGCTGTCAGAGGATCATCACTCAATTTTGAAGAGCTCCG 660
Db 909 GTGAATGACGGCTTCAAGGCTGTCAGAGGATCATCACTCAATTTTGAAGAGCTCCG 968
QY 661 ATCGAATTTCTAGACGTGAACCATTCACCATATGTGAGAGAGTCAATTTCAATGTTTC 720
Db 969 ATCGAATTTCTAGACGTGAACCATTCACCATATGTGAGAGAGTCAATTTCAATGTTTC 1028
QY 721 GTGAGAGAAACGGCTTCCGACATGCCGGAAGTGAACAGGGTCTCTGCACTTCAAG 780
Db 1029 GTGAGAGAAACGGCTTCCGACATGCCGGAAGTGAACAGGGTCTCTGCACTTCAAG 1088
QY 781 ATCAATGACTCGGTCAACTCAATGAG 807
Db 1089 ATCAATGACTCGGTCAACTCAATGAG 1115

RESULT 2
US-09-799-451-250
; Sequence 250, Application US/09799451
; Patent No. 6783969
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Xue, Aidong J.
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Ma, Yungting
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Dunrui
; APPLICANT: Yang, Yonghong
```

```
; APPLICANT: Wehrman, Tom
; APPLICANT: Ghosh, Reena
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6783969el Nucleic Acids and
; FILE REFERENCE: 803
; CURRENT APPLICATION NUMBER: US/09/799,451
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 948
; SOFTWARE: pc FL_genes Version 2.0
; SEQ ID NO 250
; LENGTH: 3981
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (297)..(1118)
US-09-799-451-250

Query Match 46.8%; Score 516.8; DB 4; Length 3981;
Best Local Similarity 99.6%; Pred. No. 3.6e-159;
Matches 518; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 575 CGGACAGTAAGATCAGTGTTCCTCCGGAATGAACGGCTTCAAGCTCTGCAAGGATCA 634
Db 1 CGGACAGTAAGATCAGTGTTCCTCCGGAATGAACGGCTTCAAGCTCTGCAAGGATCA 60
QY 635 TCCACTCAATTTTGAAGAGTCTGATCGAATTCATAGACAGTGAACATTCACATAT 694
Db 61 TCCACTCAATTTTGAAGAGTCTGATCGAATTCATAGACAGTGAACATTCACATAT 120
QY 695 GTGCAAGAGATCATTTTCAAGTGTCTGTGAGAGAAACGGCTTCCGACATGCCGGAAG 754
Db 121 GTGCAAGAGATCATTTTCAAGTGTCTGTGAGAGAAACGGCTTCCGACATGCCGGAAG 180
QY 755 TGGAACGGGCTCTGCAAGCTTCAAGATCATGACTGGTCACTCAATGAGAACGCT 814
Db 181 TGGAACGGGCTCTGCAAGCTTCAAGATCATGACTGGTCACTCAATGAGAACGCT 240
QY 815 TTTCTGTGGAAGACATTAATTAATCTGTGTCCAGGCGCTATCTTAAGAAATGGGATGA 874
Db 241 TTTCTGTGGAAGATTAATTAATCTGTGTCCAGGCGCTATCTTAAGAAATGGGATGA 300
QY 875 AAGCTGACTCTCAGTGAAGATGAACAGATGCGCTCTTTTATCTCAATGTTGTATCA 934
Db 301 AAGCTGACTCTCAGTGAAGATGAACAGATGCGCTCTTTTATCTCAATGTTGTATCA 360
QY 935 TCACCAACACACATGTTTGAAGGTTCCATCCGAGCATGCGCTGTGATCCGTTCC 994
Db 361 TCACCAACACACATGTTTGAAGGTTCCATCCGAGCATGCGCTGTGATCCGTTCC 420
QY 995 TGCTCTAGCCCTGCTCTCTGTGTGTTTGGCCCTCTGTGCACTGTGATATCA 1054
Db 421 TGCTCTAGCCCTGCTCTCTGTGTGTTTGGCCCTCTGTGCACTGTGATATCA 480
QY 1055 AGGAGTTCCTCCACCCCTGCGGAGAGAGTGAAGAAA 1094
Db 481 AGGAGTTCCTCCACCCCTGCGGAGAGAGTGAAGAAA 520

RESULT 3
US-09-774-528-297
; Sequence 297, Application US/09774528
; Patent No. 6743619
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Ren, Feiyan
; APPLICANT: Zhang, Jie
; APPLICANT: Zhao, Qing A.
```





Dp	1011	CCATAGCCTCACAAGATTAAATTCGAGTTTGAGCTGCAGGTGAGGAGGACTTGCAATYYYY	1070
Qy	749	GCAACGTGAGACAGGCTCCTCTGCAGCTTCAAGATTCATGACTCGCTCACACTCAATGAGA	808
Dp	1071	YY	1130
Qy	809	AGCCCTTTCGTGAGGAAGACCTATTACTGTGTGCAGGCGCTATCTTAAAGAAGTGG	868
Dp	1131	YY	1190
Qy	869	GCATGAAGAGTGCACTCCAGCTCAGACATGAACGATGCGCTCTTTATCTCCAGTTCTG	928
Dp	1191	YY	1250
Qy	929	TCATCATCACCACACACACTGTTCTGACGGTTCACATCTGSCATCGCCCTGCTATCC	988
Dp	1251	YY	1310
Qy	989	TGTTCTGTGCTCTAGCCCTGGCTCTCTCTGTGTGTTCTGGCCCTCTGCTGCACGTGA	1048
Dp	1311	YY	1370
Qy	1049	TTATCAAGAGAGTCCTCCACCCCTGCG	1077
Dp	1371	YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	1399

```

RESULT 5
US-09-621-976-8976/c
; Sequence 8976, Application US/09621976
; Patent NO. 663063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 8976
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-8976

```

Query Match	3.8%	Score 42.4	DB 4	Length 399
-------------	------	------------	------	------------

Matches	25;	Conservative	142;	Mismatches	113;	Indels	0;	Gaps	0;
---------	-----	--------------	------	------------	------	--------	----	------	----

[illegible]

## RESULT 6

```

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

```

Query Match 3.8%; Score 41.4; DB 1; Length 7218;

Matches 27; Conservative 198; Mismatches 174; Indels 0; Gaps 0;

Oy	626	GAGAACAACTTATATGAACGTCAGAGAACAATCCGTCAAGCGCTTAGAAG	322
Db	1407	RRR	1348
Oy	323	AACCTCCAGAAAGTTCTGCAGAGAGACTTCATGACATGAAGATTGAAAGGCCA	382
Db	1347	RRR	1288
Oy	383	GTGACGAGATTATATGAAGAAGACAGACAGGGTACAGACAGCAGCTCATATGCTT	442
Db	1287	RRR	1228
Oy	443	TGACTGATGGAAGCTCCAGACAGATCTCTTTTCTATTCAGAGGAGAGCTAATAGT	502
Db	1227	RRR	1168
Oy	503	CTGACGATCTTGAGCAATGTTTAACGTGCTGTGTGGAAGAATTCATCATGACACAGC	562
Db	1167	RRR	1108
Oy	563	TGGCCCGAGTTCCGACAGTAGAAGATCATGTGTTCCCGTAGACAGCGCTTCAGGCTC	622

 $QY$



```

; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
; US-09-806-708B-22
```

```

Query Match          3.4%; Score 37.6; DB 4; Length 1141;
Best Local Similarity 10.9%; Pred. No. 0.26; Mismatches 443; Indels 6; Gaps 1;
Matches 89; Conservative 281; Mismatches 443; Indels 6; Gaps 1;
```

```

Qy 286 ACAGAGACAGAGACAATCCGTCAGAGCCCTAGAGAACTCCAGAAAGTTCTCCAGCA 345
D 987 AAKAMAGAGNNMRYGAAAGNKGWCAAAATGMBWADTAGKCNNNNNNTTIDVARM 928
Qy 346 GGAGACACTTATGATGATGAAGATTGAAAAGGCCGTGAGAGATTATATGAAAC 405
D 927 MKAKNNNNNNNAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 868
Qy 406 AGACAGAGGTACAGAGACAGAGAGGTCATGATGATGATGATGATGATGATGATGAT 465
D 867 NARTVYHHHAAKMMMAWMTTNNNNNNNNNAACNTTNTTMMABKMSKCNNNNNNNN 808
Qy 466 GATCTCTTTTCTATGAGAGAGAGAGGCTAATAGTCTGAGATCTTGTGCAATGTT 525
D 807 NNNNTWCHYTTANABCYRANNNNAARMAATCNVYAAVTTHTDWCYKTMNTWYMD 748
Qy 526 TACTGTGTGTGTAAGAAATTTCAATGAGACAGAGTGGCCGAGATGGGAGACATA 585
D 747 MMTTMTTTRTNTTNTTNTNNNNNNNACTNNNNNNNAAYAAATNNNGCWNNTDAR 688
Qy 586 GATCATGTGTTTCCGTAATGAGAGGCTTTCAGGCTCTGCAAGCATCATCTCATCT 645
D 687 RTNTTNTTMRRTMTNTTTRTSTTRHHYTGATNNNNNNNNNNNNNNNNNNNNNNNN 628
Qy 646 TTGAAGAGTCTCCGATCGAATTTCTAGCAGCTGACATCCACCATATGTGAGAGAG 705
D 627 RMTWKGDGMYTRKKVKRDTCTVVDWADSWVMVYNNMRRCRDVYTRNTNYCKSYAHS 568
Qy 706 TCATTTCAAGTGTCTGAGAGAGAAAGGCTTCGAGATGCGCCGCAACGTGACAGGTC 765
D 567 YWYNNNNAMMYRRYSANWSSMAWTTTNNNNNNNNNNNNNNNNNNNNNNNNNTDTRY 508
Qy 766 CTCGAGCTTCAAGATCAATGATCGTCACTGATCATGATGAGAGCCCTTTCTGTGAA 825
D 507 WWWWKMAABTTTVDSCNAKSMRGNNMRAAMWMAANDAGAMDTTMMGNTTMMR 448
Qy 826 GACACTATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 885
D 447 RAMMMNNMAMCRAYCCNNNNNNRACVHKH-----MMRTWYKMMKAACNNNNNNBR 394
Qy 886 CAGGTGAGATGAACGATGAGCTCTTTTATCTCCAGTTCTGTCATCATCAACACACA 945
D 393 MRVAMMYSRJTTTIDMMWTSDBWBMWYTVDTMMMAAMNNNNNNNNNNNNNNNNNN 334
Qy 946 CACTGTCTGAGCGGTCTGATCGGCTGATCGGCTGATCGGCTGATCGGCTGATCGGCT 1005
D 333 HMAHTCTGNNNTSAYBMAAMSMMAAGASBNBYTNCMTTGTGKTMTNNNNNNNNNN 274
Qy 1006 CTGAGCTCTCTGAGGTTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCT 1065
D 273 RTKTVAMCNRNRYDYDTAVMTBKRYKYCYAVBYMYBMVGMKHBMBRABHBSMMWV 214
Qy 1066 CCACCCCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104
D 213 KCRNKYVSMYHAMRYBKMABAVGCNNNMKDRMAHHH 175
```

```

RESULT 10
US-08-728-323A-1/c
; Sequence 1, Application US/08728323A
```

```

; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MS/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ. ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
; US-08-728-323A-1
```

```

Query Match          3.4%; Score 37.4; DB 2; Length 3489;
Best Local Similarity 50.3%; Pred. No. 0.67; Mismatches 91; Indels 0; Gaps 0;
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;
```

```

Qy 883 CTCGAGTCAAGATGAGAGGCTCTCTTTATCTCAGTTCTGATCATGATCACCACC 942
D 1149 CTCCTGTCATCTCTCTGTCATCTCTGTCATCTCTGTCATCTCTGTCATCTCTGTCAT 1090
Qy 943 ACACACTGTTGACGAGTTCCATCTGAGCATGAGCCCTGCTATCTGTTCTGCTCTA 1002
D 1089 CTCCTGTCATCTCTCTGTCATCTCTGTCATCTCTGTCATCTCTGTCATCTCTGTCAT 1030
Qy 1003 GCCCTGCTCTCTCTGAGGTTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGG 1062
D 1029 CTCCTGCTCTCTCTGATCTGCTCTGCTCTGCTCTGCTCTGCTCTGATCTTATGATGAT 970
Qy 1063 CCT 1065
D 969 CTT 967
```

```

RESULT 11
US-09-298-568-1/c
; Sequence 1, Application US/09298568
; Patent No. 6323792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
```

APPLICANT: Ballestas, Mary E.  
APPLICANT: Kaye, Kenneth M.  
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
FILE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE  
FILE REFERENCE: 16412-10001R  
CURRENT APPLICATION NUMBER: US/09/298,568  
EARLIER FILING DATE: 1999-04-21  
EARLIER APPLICATION NUMBER: US 60/109,422  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 3489  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-298-568-1

Query Match 3.4%; Score 37.4; DB 3; Length 3489;  
Best Local Similarity 50.3%; Pred. No. 0.67;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

883 CTCGAGTCAGCATGAGATGCGCTCTTTATCTCCAGTTCTGTCATCATCAGCACC 942  
1149 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1090  
943 ACACACTGTTCTGACGGTTCCATCTGCGCATCGCCCTGTCATCTGTTCTGCTCCTA 1002  
1089 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1030  
1003 GCCCTGCTCTCTCTGTCATCTCTGCGCCCTCTGTCATCTGTCATCTGTCATCTGTCATC 1062  
1029 CTCCTCGATCCTCTCTGTCATCTCTCTGTCATCTCTCTCGATCCTCTCTGTCATC 970

Qy 1063 CCT 1065  
Db 969 CTT 967

RESULT 12  
US-09-410-399-1/c  
Sequence 1, Application US/09410399  
Patent No. 6482587  
GENERAL INFORMATION:  
APPLICANT: Robertson, Erle S.  
APPLICANT: Cocter, Murray A.  
TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA  
FILE REFERENCE: UM-03778  
CURRENT APPLICATION NUMBER: US/09/410,399  
EARLIER FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 3489  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-410-399-1

Query Match 3.4%; Score 37.4; DB 4; Length 3489;  
Best Local Similarity 50.3%; Pred. No. 0.67;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

883 CTCGAGTCAGCATGAGATGCGCTCTTTATCTCCAGTTCTGTCATCATCAGCACC 942  
1149 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1090  
943 ACACACTGTTCTGACGGTTCCATCTGCGCATCGCCCTGTCATCTGTTCTGCTCCTA 1002  
1089 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1030  
1003 GCCCTGCTCTCTCTGTCATCTCTGCGCCCTCTGTCATCTGTCATCTGTCATCTGTCATC 1062  
1029 CTCCTCGATCCTCTCTGTCATCTCTCTGTCATCTCTCTCGATCCTCTCTGTCATC 970

Qy 1063 CCT 1065  
Db 969 CTT 967

RESULT 13  
US-09-894-273-1/c  
Sequence 1, Application US/09894273  
Patent No. 6756203  
GENERAL INFORMATION:  
APPLICANT: Kieff, Elliott D.  
APPLICANT: Ballestas, Mary E.  
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
FILE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE  
FILE REFERENCE: 16412-10001R  
CURRENT APPLICATION NUMBER: US/09/894,273  
EARLIER FILING DATE: 2001-06-28  
PRIOR APPLICATION NUMBER: US 60/109,422  
PRIOR FILING DATE: 1998-11-19  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
LENGTH: 3489  
TYPE: DNA  
ORGANISM: Kaposi's sarcoma-associated herpesvirus  
US-09-894-273-1

Query Match 3.4%; Score 37.4; DB 4; Length 3489;  
Best Local Similarity 50.3%; Pred. No. 0.67;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

883 CTCGAGTCAGCATGAGATGCGCTCTTTATCTCCAGTTCTGTCATCATCAGCACC 942  
1149 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1090  
943 ACACACTGTTCTGACGGTTCCATCTGCGCATCGCCCTGTCATCTGTTCTGCTCCTA 1002  
1089 CTCCTCGATCCTCTCTGTCATCTCTCTCGATCCTCTCTCGATCCTCTCTGTCATC 1030  
1003 GCCCTGCTCTCTCTGTCATCTCTGCGCCCTCTGTCATCTGTCATCTGTCATCTGTCATC 1062  
1029 CTCCTCGATCCTCTCTGTCATCTCTCTGTCATCTCTCTCGATCCTCTCTGTCATC 970

Qy 1063 CCT 1065  
Db 969 CTT 967

RESULT 14  
US-08-770-379-20  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Ruoso, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
FILE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770.379  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 3.4%; Score 37.4; DB 2; Length 32207;  
Best Local Similarity 50.3%; Pred. No. 3.5;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 883 CTCGAGTCAAGATGAGGCTCTCTTTATCTCCAGTTCTGTCATCACCACC 942  
DB 20848 CTCCTCGTCATCTCTGTCATCTCTCGTCATCTCTCGTCATCTCTGTCATC 20907  
QY 943 ACACACTGTTGACGCTTCATCTGCGCCATGCGCTGTCATCTCTGTCATC 1002  
DB 20908 CTCCTCGTCATCTCTGTCATCTCTCGTCATCTCTCGTCATCTCTGTCATC 20967  
QY 1003 GCCCTGCTCTCTCTGTCATCTGCGCCATGCGCTGTCATCTGTCATC 1062  
DB 20968 CTCCTGCTCTCTCTGTCATCTGTCATCTCTGTCATCTCTGTCATC 21027  
QY 1063 CCT 1065  
DB 21028 CTT 21030

RESULT 15  
US-08-757-669A-20  
Sequence 20, Application US/08757669A  
Patent No. 6183751  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,669A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 45185-F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-757-669A-20

Query Match 3.4%; Score 37.4; DB 3; Length 32207;  
Best Local Similarity 50.3%; Pred. No. 3.5;  
Matches 92; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 883 CTCGAGTCAAGATGAGGCTCTCTTTATCTCCAGTTCTGTCATCACCACC 942  
DB 20848 CTCCTCGTCATCTCTGTCATCTCTCGTCATCTCTCGTCATCTCTGTCATC 20907  
QY 943 ACACACTGTTGACGCTTCATCTGCGCCATGCGCTGTCATCTCTGTCATC 1002  
DB 20908 CTCCTCGTCATCTCTGTCATCTCTCGTCATCTCTCGTCATCTCTGTCATC 20967  
QY 1003 GCCCTGCTCTCTCTGTCATCTGCGCCATGCGCTGTCATCTGTCATC 1062  
DB 20968 CTCCTGCTCTCTCTGTCATCTGTCATCTCTGTCATCTCTGTCATC 21027  
QY 1063 CCT 1065  
DB 21028 CTT 21030

Search completed: June 14, 2005, 16:47:10  
Job time : 233 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: June 14, 2005, 15:32:23 ; Search time 783 Seconds  
(without alignments)  
8740.438 Million cell updates/sec

Title: US-09-970-076-1\_COPY\_104\_1207  
Perfect score: 1104  
Sequence: 1 atggccacggcgagcgag.....gtcgaagaaataataaaaa 1104

Scoring table:  
IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6046767 seqs, 3099530249 residues

Total number of hits satisfying chosen parameters: 12093534

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*
- 20: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*
- 21: /cgn2\_6/ptodata/2/pubpna/US10J\_PUBCOMB.seq:\*
- 22: /cgn2\_6/ptodata/2/pubpna/US10K\_PUBCOMB.seq:\*
- 23: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq:\*
- 24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*
- 25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1104	100.0	1454	17	US-10-133-937-58
2	1104	100.0	1454	17	US-10-159-563-58
3	1090.8	98.8	5540	10	US-09-918-715-176
4	1090.8	98.8	5540	10	US-09-918-715-231
5	1090.8	98.8	5540	15	US-10-301-822-198
6	1090.8	98.8	5540	20	US-10-474-794-176
7	1090.8	98.8	5540	20	US-10-474-794-231

8	960.8	87.0	1674	15	US-10-038-307-17	Sequence 17, Appl
9	960.8	87.0	1674	15	US-10-201-292-17	Sequence 17, Appl
10	952.4	86.3	1650	15	US-10-038-307-13	Sequence 13, Appl
11	952.4	86.3	1650	15	US-10-038-307-15	Sequence 15, Appl
12	952.4	86.3	1650	15	US-10-201-292-13	Sequence 13, Appl
13	952.4	86.3	1650	15	US-10-201-292-15	Sequence 15, Appl
14	950.8	86.1	1650	15	US-10-038-307-9	Sequence 9, Appl
15	950.8	86.1	1650	15	US-10-201-292-9	Sequence 9, Appl
16	949.4	86.0	1056	15	US-10-038-307-23	Sequence 23, Appl
17	949.4	86.0	1056	15	US-10-201-292-23	Sequence 23, Appl
18	949.4	86.0	1713	15	US-10-038-307-19	Sequence 19, Appl
19	949.4	86.0	1713	15	US-10-201-292-19	Sequence 19, Appl
20	949.4	86.0	2272	15	US-09-796-753-11	Sequence 11, Appl
21	949.4	86.0	2272	15	US-10-038-307-1	Sequence 1, Appl
22	949.4	86.0	2272	15	US-10-201-292-1	Sequence 1, Appl
23	949.4	86.0	2353	14	US-10-198-846-9957	Sequence 957, Ap
24	949.4	86.0	1008	15	US-10-038-307-25	Sequence 25, Appl
25	949.4	86.0	1008	15	US-10-201-292-25	Sequence 25, Appl
26	912.4	82.6	1047	15	US-10-038-307-21	Sequence 21, Appl
27	912.4	82.6	1047	15	US-10-201-292-21	Sequence 21, Appl
28	892.6	80.9	1608	15	US-10-201-292-35	Sequence 35, Appl
29	867.6	78.6	1623	15	US-10-038-307-11	Sequence 11, Appl
30	867.6	78.6	1623	15	US-10-201-292-11	Sequence 11, Appl
31	860	77.9	2397	17	US-10-062-674-1757	Sequence 1757, Ap
32	853.4	77.3	5220	10	US-09-918-715-186	Sequence 186, App
33	853.4	77.3	5220	10	US-09-918-715-300	Sequence 300, App
34	853.4	77.3	5220	20	US-10-474-794-186	Sequence 186, App
35	853.4	77.3	5220	20	US-10-474-794-300	Sequence 300, App
36	819	74.2	1534	15	US-10-201-292-33	Sequence 33, Appl
37	803.8	72.8	1609	15	US-10-037-270-8	Sequence 8, Appl
38	803.8	72.8	1609	17	US-10-117-722-8	Sequence 8, Appl
39	802.2	72.7	1718	20	US-10-357-930-30300	Sequence 30300, A
40	750	67.9	1464	15	US-10-201-292-31	Sequence 31, Appl
41	692	62.7	1401	15	US-10-201-292-27	Sequence 27, Appl
42	692	62.7	1401	15	US-10-201-292-29	Sequence 29, Appl
43	521	47.2	1650	13	US-10-047-542-98	Sequence 98, Appl
44	521	47.2	1650	13	US-10-047-542-100	Sequence 100, App
45	516.8	46.8	3981	18	US-10-302-172-250	Sequence 250, App

#### ALIGNMENTS

RESULT 1  
US-10-133-937-58  
Sequence 58, Application US/10133937  
Publication No. US0030207278A1  
GENERAL INFORMATION:  
APPLICANT: Khan, Javed  
APPLICANT: Ringner, Markus  
APPLICANT: Peterson, Carsten  
APPLICANT: Melitzer, Paul  
TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING, DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND  
FILE REFERENCE: 11613.56US01  
CURRENT APPLICATION NUMBER: US/10/133,937  
NUMBER OF SEQ ID NOS: 99  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 58  
LENGTH: 1454  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-133-937-58

Query Match 100.0%; Score 1104; DB 17; Length 1454;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 144 ATGGCCACGGCGAGCGAGAGAGCCCTTCGCGATCGCTTCAGTGGCTCTTTGGCCACT 60  
1 ATGGCCACGGCGAGCGAGAGAGCCCTTCGCGATCGCTTCAGTGGCTCTTTGGCCACT 60  
144 ATGGCCACGGCGAGCGAGAGAGCCCTTCGCGATCGCTTCAGTGGCTCTTTGGCCACT 203

QY 61 CTGGTCTCATCTGCGCGGCAAGGGGAGCGCAGGAGGATGGGGGTCCAGCCTGTAC 120  
DB 204 CTGGTCTCATCTGCGCGGCAAGGGGAGCGCAGGAGGATGGGGGTCCAGCCTGTAC 263  
QY 121 GGGCGATTTGACCTGTACTTCTATTTTGGACAAATCAGAAAGTGTCTGCACCTGAAT 180  
DB 264 GGGCGATTTGACCTGTACTTCTATTTTGGACAAATCAGAAAGTGTCTGCACCTGAAT 323  
QY 181 GAAATCTATTACTTGTGAGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 240  
DB 324 GAAATCTATTACTTGTGAGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 383  
QY 241 TCCCTTATTTGTTTTCTCACCCGAGAAACAATTGAAATGAACTGACAGAAACAGAGAA 300  
DB 384 TCCCTTATTTGTTTTCTCACCCGAGAAACAATTGAAATGAACTGACAGAAACAGAGAA 443  
QY 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACATTACATG 360  
DB 444 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACATTACATG 503  
QY 361 CATGAAGATTTGAAAGGGCCAGTAGAGACATTTATTATGAAAAACAGACAAGGGTACAG 420  
DB 504 CATGAAGATTTGAAAGGGCCAGTAGAGACATTTATTATGAAAAACAGACAAGGGTACAG 563  
QY 421 ACAGCCAGCGTCATCATTTGCTTTGATGATGAGAGACCTCAGTAAGATCTTTTCTAT 480  
DB 564 ACAGCCAGCGTCATCATTTGCTTTGATGATGAGAGACCTCAGTAAGATCTTTTCTAT 623  
QY 481 TCAGAGAGGAGGCTTAATAGGTCTGAGAGATTTGGTGCATTTGTTTCTGTGTG 540  
DB 624 TCAGAGAGGAGGCTTAATAGGTCTGAGAGATTTGGTGCATTTGTTTCTGTGTG 683  
QY 541 AAGATTTCAATGAGACACAGCTGGCCGGATTGCGGACAGTAAAGATCATGTGTTTCCC 600  
DB 684 AAGATTTCAATGAGACACAGCTGGCCGGATTGCGGACAGTAAAGATCATGTGTTTCCC 743  
QY 601 GTGAATGACGGCTTTGAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTGTC 660  
DB 744 GTGAATGACGGCTTTGAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTGTC 803  
QY 661 ATCGAAATTTCTAGAGCTGAACCATCCACCTATGTGCAAGAGAGTCAATTTCAAGTTGTC 720  
DB 804 ATCGAAATTTCTAGAGCTGAACCATCCACCTATGTGCAAGAGAGTCAATTTCAAGTTGTC 863  
QY 721 GTGAGAGAAAAGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAAG 780  
DB 864 GTGAGAGAAAAGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGACGCTTCAAG 923  
QY 781 ATCAATGACTCGGTCAACCTCAATGAGAAAGCCCTTTTCTGTGAAACACTTATTTTACTG 840  
DB 924 ATCAATGACTCGGTCAACCTCAATGAGAAAGCCCTTTTCTGTGAAACACTTATTTACTG 983  
QY 841 TGTCTCAGGCGCTATCTTAAAAAGAGTTGGCATGAAAGCTGCACTCCAGGTCAAGTAAAC 900  
DB 984 TGTCTCAGGCGCTATCTTAAAAAGAGTTGGCATGAAAGCTGCACTCCAGGTCAAGTAAAC 1043  
QY 901 GATGAGCGCTCTTTTATCTCCAGTTCTGTCAATCATCAACCAACACTGTTCTGAGGAT 960  
DB 1044 GATGAGCGCTCTTTTATCTCCAGTTCTGTCAATCATCAACCAACACTGTTCTGAGGAT 1103  
QY 961 TCCATCTGGGCATCGCGCTGCTGATCTGTCTCTGCTCTGAGCCCTGCTCTCTGAG 1020  
DB 1104 TCCATCTGGGCATCGCGCTGCTGATCTGTCTCTGCTCTGAGCCCTGCTCTCTGAG 1163  
QY 1021 TGGTTCCTGAGCCCTCTGTGCACTGTGATTAACAAGAGGTCTCTCAACCCCTGCGGAG 1080  
DB 1164 TGGTTCCTGAGCCCTCTGTGCACTGTGATTAACAAGAGGTCTCTCAACCCCTGCGGAG 1223  
QY 1081 GAGAGTGAGGAAATTAATAATAAA 1104  
DB 1224 GAGAGTGAGGAAATTAATAATAAA 1247

RESULT 2  
US-10-159-563-58  
; Sequence 58, Application US/10159563  
; Publication No. US20040009154A1  
; GENERAL INFORMATION:  
; APPLICANT: Khan, Javed  
; APPLICANT: Ringner, Markus  
; APPLICANT: Peterson, Carsten  
; APPLICANT: Melzer, Paul  
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR  
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS  
; FILE REFERENCE: 11613.56US11  
; CURRENT APPLICATION NUMBER: US/10/159,563  
; PRIOR FILING DATE: 2002-12-09  
; PRIOR APPLICATION NUMBER: US 10/133,937  
; NUMBER OF SEQ ID NOS: 444  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1454  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-159-563-58  
  
Query Match 100.0%; Score 1104; DB 17; Length 1454;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATGGCCACGGCGGAGGAGAGACCCCTCGGCATCGGCTTCAAGTGGCTCTTTGGCCACT 60  
DB 144 ATGGCCACGGCGGAGGAGAGACCCCTCGGCATCGGCTTCAAGTGGCTCTTTGGCCACT 203  
QY 61 CTGGTCTCATCTGCGCGGCAAGGGGAGCGCAGGAGGATGGGGGTCCAGCCTGTAC 120  
DB 204 CTGGTCTCATCTGCGCGGCAAGGGGAGCGCAGGAGGATGGGGGTCCAGCCTGTAC 263  
QY 121 GGGCGATTTGACCTGTACTTCTATTTTGGACAAATCAGAAAGTGTCTGCACCTGAAT 180  
DB 264 GGGCGATTTGACCTGTACTTCTATTTTGGACAAATCAGAAAGTGTCTGCACCTGAAT 323  
QY 181 GAAATCTATTACTTGTGAGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 240  
DB 324 GAAATCTATTACTTGTGAGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 383  
QY 241 TCCCTTATTTGTTTTCTCACCCGAGAAACAATTGAAATGAACTGACAGAAACAGAGAA 300  
DB 384 TCCCTTATTTGTTTTCTCACCCGAGAAACAATTGAAATGAACTGACAGAAACAGAGAA 443  
QY 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACATTACATG 360  
DB 444 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACATTACATG 503  
QY 361 CATGAAGATTTGAAAGGGCCAGTAGAGACATTTATTATGAAAAACAGACAAGGGTACAG 420  
DB 504 CATGAAGATTTGAAAGGGCCAGTAGAGACATTTATTATGAAAAACAGACAAGGGTACAG 563  
QY 421 ACAGCCAGCGTCATCATTTGCTTTGATGATGAGAGACCTCAGTAAGATCTTTTCTAT 480  
DB 564 ACAGCCAGCGTCATCATTTGCTTTGATGATGAGAGACCTCAGTAAGATCTTTTCTAT 623  
QY 481 TCAGAGAGGAGGCTTAATAGGTCTGAGAGATTTGGTGCATTTGTTTCTGTGTG 540  
DB 624 TCAGAGAGGAGGCTTAATAGGTCTGAGAGATTTGGTGCATTTGTTTCTGTGTG 683  
QY 541 AAGATTTCAATGAGACACAGCTGGCCGGATTGCGGACAGTAAAGATCATGTGTTTCCC 600  
DB 684 AAGATTTCAATGAGACACAGCTGGCCGGATTGCGGACAGTAAAGATCATGTGTTTCCC 743  
QY 601 GTGAATGACGGCTTTGAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTGTC 660  
DB 744 GTGAATGACGGCTTTGAGGCTCTGCAAGGACATCATCTCAATTTTGAAGAGTCTGTC 803



Qy 661 ATCAAAATTTAGAGCTGAAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
 Db 804 ATCAAAATTTAGAGCTGAAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 863  
 Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGCTCTGCAAGTTCAAG 780  
 Db 864 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGCTCTGCAAGTTCAAG 923  
 Qy 781 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTCTGTGAGAACACTTATTACTG 840  
 Db 924 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTCTGTGAGAACACTTATTACTG 983  
 Qy 841 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTTCAGATGAA 900  
 Db 984 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTTCAGATGAA 1043  
 Qy 901 GATGGCTCTCTTTATCTCAAGTTCTGATTCATCATCAACCAACACTGTTCTGAAGGT 960  
 Db 1044 GATGGCTCTCTTTATCTCAAGTTCTGATTCATCATCAACCAACACTGTTCTGAAGGT 1103  
 Qy 961 TCCATCTGCGCAATCGCCCTGCTGATCTGTTCTGCTAGCCCTGCTCTCTG 1020  
 Db 1104 TCCATCTGCGCAATCGCCCTGCTGATCTGTTCTGCTAGCCCTGCTCTCTG 1163  
 Qy 1021 TGGTTCTGCGCCCTCTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCGAG 1080  
 Db 1164 TGGTTCTGCGCCCTCTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCGAG 1223  
 Qy 1081 GAGAGTGAGGAAATTAATAATTA 1104  
 Db 1224 GAGAGTGAGGAAATTAATAATTA 1247

RESULT 3

US-09-918-715-176  
 / Sequence 176, Application US/09918715  
 / Publication No. US20030017157A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Brad St. Croix  
 / APPLICANT: Bert Vogelstein  
 / APPLICANT: Kenneth Kinzler  
 / TITLE OF INVENTION: ENDOTHELIAL CBEL EXPRESSION PATTERNS  
 / FILE REFERENCE: 1107.00134  
 / CURRENT APPLICATION NUMBER: US/09/918, 715  
 / PRIOR FILING DATE: 2001-08-01  
 / PRIOR APPLICATION NUMBER: 60/222,599  
 / PRIOR FILING DATE: 2000-08-02  
 / PRIOR APPLICATION NUMBER: 60/224,360  
 / PRIOR FILING DATE: 2000-08-11  
 / PRIOR APPLICATION NUMBER: 60/282,850  
 / PRIOR FILING DATE: 2000-04-11  
 / NUMBER OF SEQ ID NOS: 358  
 / SOFTWARE: FastSeq for Windows Version 3.0  
 / SEQ ID NO 176  
 / LENGTH: 5540  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / US-09-918-715-176

Query Match 98.8%; Score 1090.8; DB 10; Length 5540;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCACGGCGAGGAGGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTCTTGGCCACT 60  
 Db 144 ATGGCCACGGCGAGGAGGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTCTTGGCCACT 203  
 Qy 61 CTGGTGTCTATCTGCGCGGCAAGGGGAGCGAGGAGAGTGGGGTCCAGCTGTCTAC 120  
 Db 204 CTGGTGTCTATCTGCGCGGCAAGGGGAGCGAGGAGAGTGGGGTCCAGCTGTCTAC 263  
 Qy 121 GGGGATTTTACCTGTAATTTTGAAGAAATCAGGAAGTGTCTGACCACTGGAAAT 180  
 Db 122 GGGGATTTTACCTGTAATTTTGAAGAAATCAGGAAGTGTCTGACCACTGGAAAT 180

Db 264 GGGGATTTTACCTGTAATTTTGAAGAAATCAGGAAGTGTCTGACCACTGGAAAT 323  
 Qy 181 GAAATCTATTACTTTTGGAAACAGTTGGCTCACAAAATTCATCAGCCACAGTTGAGATG 240  
 Db 324 GAAATCTATTACTTTTGGAAACAGTTGGCTCACAAAATTCATCAGCCACAGTTGAGATG 383  
 Qy 241 TCCTTTATTTTCTTCACCCGAGAAACCTTAAATGAACTGACAGAAACAGAA 300  
 Db 384 TCCTTTATTTTCTTCACCCGAGAAACCTTAAATGAACTGACAGAAACAGAA 443  
 Qy 301 CAATTCGTAAAGCCCTGAGAAAGATTCAGAAAGTTCTGACAGAGAGACACTTAATG 360  
 Db 444 CAATTCGTAAAGCCCTGAGAAAGATTCAGAAAGTTCTGACAGAGAGACACTTAATG 503  
 Qy 361 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAACAGACAGAGGTACAG 420  
 Db 504 CATGAAGATTTTGAAGGGCCAGTGAAGATTTATATGAAACAGACAGAGGTACAG 563  
 Qy 421 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAAGATCTTTTCTAT 480  
 Db 564 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAAGATCTTTTCTAT 623  
 Qy 481 TCAAGAGGAGGCTTAATGAGTCTCGAGATCTTGTGCAATTTGTTACTGTGTGTG 540  
 Db 624 TCAAGAGGAGGCTTAATGAGTCTCGAGATCTTGTGCAATTTGTTACTGTGTGTG 683  
 Qy 541 AAAGATTTTCAATGAGACAGAGTGGCCGGAATTCGGAAGTAAGATGATTTTCC 600  
 Db 684 AAAGATTTTCAATGAGACAGAGTGGCCGGAATTCGGAAGTAAGATGATTTTCC 743  
 Qy 601 GTGAATGACGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTG 660  
 Db 744 GTGAATGACGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTG 803  
 Qy 661 ATCAAAATTTAGAGCTGAAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
 Db 804 ATCAAAATTTAGAGCTGAAACCATCCATATGTGAGAGAGTCAATTTCAAGTTGTC 863  
 Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGCTCTGCAAGTTCAAG 780  
 Db 864 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGCTCTGCAAGTTCAAG 923  
 Qy 781 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTCTGTGAGAACACTTATTACTG 840  
 Db 924 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTCTGTGAGAACACTTATTACTG 983  
 Qy 841 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTTCAGATGAA 900  
 Db 984 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTTCAGATGAA 1043  
 Qy 901 GATGGCTCTCTTTATCTCAAGTTCTGATTCATCATCAACCAACACTGTTCTGAAGGT 960  
 Db 1044 GATGGCTCTCTTTATCTCAAGTTCTGATTCATCATCAACCAACACTGTTCTGAAGGT 1103  
 Qy 961 TCCATCTGCGCAATCGCCCTGCTGATCTGTTCTGCTAGCCCTGCTCTCTG 1020  
 Db 1104 TCCATCTGCGCAATCGCCCTGCTGATCTGTTCTGCTAGCCCTGCTCTCTG 1163  
 Qy 1021 TGGTTCTGCGCCCTCTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCGAG 1080  
 Db 1164 TGGTTCTGCGCCCTCTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCGAG 1223  
 Qy 1081 GAGAGTGAGGAA 1094  
 Db 1224 GAGAGTGAGGAA 1237

RESULT 4

US-09-918-715-231  
 / Sequence 231, Application US/09918715  
 / Publication No. US20030017157A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Brad St. Croix

APPLICANT: Bert Vogelstein  
APPLICANT: Kenneth Kinzler  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00134  
CURRENT APPLICATION NUMBER: US/09/918,715  
CURRENT FILING DATE: 2001-08-01  
PRIOR APPLICATION NUMBER: 60/222,599  
PRIOR FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: 60/224,360  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/282,850  
PRIOR FILING DATE: 2000-04-11  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 231  
LENGTH: 5540  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-715-231

Query Match 98.8%; Score 1090.8; DB 10; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTGGCCACT 60  
DB 144 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTGGCCACT 203  
QY 61 CTGGTCTCATCTGCGCGCGGCGGAGCGGAGAGTGGGGGTCCAGCGCTGTAC 120  
DB 204 CTGGTCTCATCTGCGCGCGGCGGAGCGGAGAGTGGGGGTCCAGCGCTGTAC 263  
QY 121 GCGCGATTTGACCTGTACTTCTTGAACAATCAGAAAGTGTCTGACCACTGGAAAT 180  
DB 264 GCGCGATTTGACCTGTACTTCTTGAACAATCAGAAAGTGTCTGACCACTGGAAAT 323  
QY 181 GAAATCTATTACTTTGTGGAACAGTGGCTCAACAATTCATCAGCCCAAGTTGAAGATG 240  
DB 324 GAAATCTATTACTTTGTGGAACAGTGGCTCAACAATTCATCAGCCCAAGTTGAAGATG 383  
QY 241 TCCCTTATTTGTTTCTCCACCGGAGAACCTTATGAAACAGAGAAACAGAGAA 300  
DB 384 TCCCTTATTTGTTTCTCCACCGGAGAACCTTATGAAACAGAGAAACAGAGAA 443  
QY 301 CAAATCCGTCAGGCTTGAAGAATCCAGAAAGTTCTGCGAGGAGACACTTACATG 360  
DB 444 CAAATCCGTCAGGCTTGAAGAATCCAGAAAGTTCTGCGAGGAGACACTTACATG 503  
QY 361 CATGAAGATTGAAAGGCGCAGTGAGACAGTTTATGAAACAGACAAAGGTTACAG 420  
DB 504 CATGAAGATTGAAAGGCGCAGTGAGACAGTTTATGAAACAGACAAAGGTTACAG 563  
QY 421 ACAGCCAGCGTCATCTTGTGATGAGTGAAGTCCATGAAAGTCTTTTCTAT 480  
DB 564 ACAGCCAGCGTCATCTTGTGATGAGTGAAGTCCATGAAAGTCTTTTCTAT 623  
QY 481 TCAGAGAGGAGGCTTAATAGGCTTCAGAGTCTTGTGCAATTTGTTACTGTGTG 540  
DB 624 TCAGAGAGGAGGCTTAATAGGCTTCAGAGTCTTGTGCAATTTGTTACTGTGTG 683  
QY 541 AAAGATTCAATGAGACACAGCTGGCCCGGATTCGCGACAGTAAAGTGTCTTCCC 600  
DB 684 AAAGATTCAATGAGACACAGCTGGCCCGGATTCGCGACAGTAAAGTGTCTTCCC 743  
QY 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGCGATCCATCTCAATTTGAAGAAGTCTG 660  
DB 744 GTGAATGACGGCTTTCAGGCTCTGCAAGGCGATCCATCTCAATTTGAAGAAGTCTG 803  
QY 661 ATCGAAATTTAGAGAGTGAACATCCACATATGTGAGAGAGATCAATTTCAAGTTGTC 720  
DB 804 ATCGAAATTTAGAGAGTGAACATCCACATATGTGAGAGAGATCAATTTCAAGTTGTC 863  
QY 721 GTGAGAGGAAAGCGCTTCCAGATGCGCGCAACGTGACAGAGTCTCTGACGTTCAAG 780

DB 864 GTGAGAGGAAAGCGCTTCCAGATGCGCGCAACGTGACAGAGTCTCTGACGTTCAAG 923  
QY 781 ATCAATGACTCGGTCACATCAATGAGAGCCCTTTCTGTGGAACATTAATTACTG 840  
DB 924 ATCAATGACTCGGTCACATCAATGAGAGCCCTTTCTGTGGAACATTAATTACTG 983  
QY 841 TGTCCAGCCCTTATCTTAAAGAAAGTTGGCATGAAAGTGCACCTCAGGTGAGATGAA 900  
DB 984 TGTCCAGCCCTTATCTTAAAGAAAGTTGGCATGAAAGTGCACCTCAGGTGAGATGAA 1043  
QY 901 GATGGCTCTCTTTATCTCCAGTTCTGTCAATCATACACACACATGTTTGAAGGT 960  
DB 1044 GATGGCTCTCTTTATCTCCAGTTCTGTCAATCATACACACACATGTTTGAAGGT 1103  
QY 961 TCCATCTGAGCATCGCCCTGCTGATCTGTCTCTGCTCTGAGCCCTGCTCTCTG 1020  
DB 1104 TCCATCTGAGCATCGCCCTGCTGATCTGTCTCTGCTCTGAGCCCTGCTCTCTG 1163  
QY 1021 TGGTTTGGCCCTCTGTGTGACCTGTGATTTATCAAGAGTCTTCCACCCCTGCGAG 1080  
DB 1164 TGGTTTGGCCCTCTGTGTGACCTGTGATTTATCAAGAGTCTTCCACCCCTGCGAG 1223  
QY 1081 GAGAGTGAAGAAA 1094  
DB 1224 GAGAGTGAAGAAA 1237

## RESULT 5

US-10-301-822-198  
Sequence 198, Application US/10301822  
Publication No. US20030148410A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
APPLICANT: Burgart, Lawrence J.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
TREATMENT OF COLON CANCER  
TITLE OF INVENTION: THERAPY OF COLON CANCER  
FILE REFERENCE: MEMO1-029P2RM  
CURRENT APPLICATION NUMBER: US/10/301,822  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 60/339,971  
PRIOR FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: US 60/361,978  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/381,988  
PRIOR FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 198  
LENGTH: 5540  
TYPE: DNA  
ORGANISM: Homo Sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (144) ... (1838)  
US-10-301-822-198

Query Match 98.8%; Score 1090.8; DB 15; Length 5540;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTGGCCACT 60  
DB 144 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTGGCCACT 203  
QY 61 CTGGTCTCATCTGCGCGCGGCGGAGCGGAGAGTGGGGGTCCAGCGCTGTAC 120

```
Db 204 CTGGTGTCTACATCGCCGCGGCAAGGCGAGCGAGGAGGATGGGGGCTTCAGCCTGCTAC 263
Qy 121 GCGGATTTGACCTGACTTCAATTTGAGCAAAATCAGAAAGTGTGACCACTGGAAT 180
Db 264 GCGGATTTGACCTGACTTCAATTTGAGCAAAATCAGAAAGTGTGACCACTGGAAT 323
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCACAGTTGAGAATG 240
Db 324 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCACAGTTGAGAATG 383
Qy 241 TCCCTTAATGTTTCTCACCAGGAAACAACTTAATGAACTGACAGAAAGAGAA 300
Db 384 TCCCTTAATGTTTCTCACCAGGAAACAACTTAATGAACTGACAGAAAGAGAA 443
Qy 301 CAAATCCGTCAAGGCTTGAAGAACTCAGAAAGTTCTGCGAGAGAGACACTTAATG 360
Db 444 CAAATCCGTCAAGGCTTGAAGAACTCAGAAAGTTCTGCGAGAGAGACACTTAATG 503
Qy 361 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAAGAGAGAGTACAGG 420
Db 504 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAAGAGAGTACAGG 563
Qy 421 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTCTAT 480
Db 564 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTCTAT 623
Qy 481 TCAGAGGGAGGCTATAGGCTCGAGATCTTGATGATGTTTCTGTTTGGTGG 540
Db 624 TCAGAGGGAGGCTATAGGCTCGAGATCTTGATGATGTTTCTGTTTGGTGG 683
Qy 541 AAGATTTCAATGAGACAGAGCTGGCCGAGATTGCGAGAGTAAAGATCAATGTTTCCC 600
Db 684 AAGATTTCAATGAGACAGAGCTGGCCGAGATTGCGAGAGTAAAGATCAATGTTTCCC 743
Qy 601 GTGAATGAGGCTTTCAAGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTGTC 660
Db 744 GTGAATGAGGCTTTCAAGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTGTC 803
Qy 661 ATCGAAATTTAGAGAGTGAACCATCAACATATGTCAGAGAGTCAATTTCAAGTTGTC 720
Db 804 ATCGAAATTTAGAGAGTGAACCATCAACATATGTCAGAGAGTCAATTTCAAGTTGTC 863
Qy 721 GTGAGAGAAACGGCTTCCGACATGCCCCGAACTGAGCAGGCTCTCTGCAAGTTCAAG 780
Db 864 GTGAGAGAAACGGCTTCCGACATGCCCCGAACTGAGCAGGCTCTCTGCAAGTTCAAG 923
Qy 781 ATCAATGACTCGGTCACTCAATGAGAGGCTTTTCTGTGAGAACTTAATTTACTG 840
Db 924 ATCAATGACTCGGTCACTCAATGAGAGGCTTTTCTGTGAGAACTTAATTTACTG 983
Qy 841 TGTCCAGGGCTTATCTTAAAGAAAGTGGATGAAAGCTCACTCCAGGTCAAGATGAAAC 900
Db 984 TGTCCAGGGCTTATCTTAAAGAAAGTGGATGAAAGCTCACTCCAGGTCAAGATGAAAC 1043
Qy 901 GATGAGCTCTTTTATCTCAAGTTCTGTCAATCATCAACCACTGTTTGAAGGCT 960
Db 1044 GATGAGCTCTTTTATCTCAAGTTCTGTCAATCATCAACCACTGTTTGAAGGCT 1103
Qy 961 TCATCTCTGAGCATGCGCTGCTGATCTGTCTCTGCTAGCCCTGAGCTTCTCTG 1020
Db 1104 TCATCTCTGAGCATGCGCTGCTGATCTGTCTCTGCTAGCCCTGAGCTTCTCTG 1163
Qy 1021 TGGTTTCTGAGGCTCTGCTGCTCACTGTGATTAATCAAGAGAGTCCCTCAACCCCTGCGAG 1080
Db 1164 TGGTTTCTGAGGCTCTGCTGCTCACTGTGATTAATCAAGAGAGTCCCTCAACCCCTGCGAG 1223
Qy 1081 GAGAGTGAAGAAA 1094
Db 1224 GAGAGTGAAGAAA 1237
```

RESULT 6

```
US-10-474-794-176
; Sequence 176, Application US/10474794
; Publication No. US20040213793A1
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107, 00179
; CURRENT APPLICATION NUMBER: US/10/474,794
; PRIORITY FILING DATE: 2003-10-14
; PRIORITY FILING DATE: 2001-04-11
; PRIORITY FILING DATE: 2001-04-11
; PRIORITY FILING DATE: 2001-08-01
; NUMBER OF SEQ. ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-474-794-176

Query Match 98.8%; Score 1090.8; DB 20; Length 5540;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCACGGCGGAGGCGAGAGCCCTGCGCATGGCTTCCAGTGGCTCTTTGGCCACT 60
Db 144 ATGGCCACGGCGGAGGCGAGAGCCCTGCGCATGGCTTCCAGTGGCTCTTTGGCCACT 203
Qy 61 CTGGTGTCTACATCGCCGCGGCAAGGCGAGCGAGGAGATGGGGGTCCAGCTGCTAC 120
Db 204 CTGGTGTCTACATCGCCGCGGCAAGGCGAGCGAGGAGATGGGGGTCCAGCTGCTAC 263
Qy 121 GCGGATTTGACCTGACTTCAATTTGAGCAAAATCAGAAAGTGTGACCACTGGAAT 180
Db 264 GCGGATTTGACCTGACTTCAATTTGAGCAAAATCAGAAAGTGTGACCACTGGAAT 323
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCACAGTTGAGAATG 240
Db 324 GAAATCTATTACTTTGTGGAACAGTTGGCTCAGAAATTCATCAGCCACAGTTGAGAATG 383
Qy 241 TCCCTTAATGTTTCTCACCAGGAAACAACTTAATGAACTGACAGAAAGAGAA 300
Db 384 TCCCTTAATGTTTCTCACCAGGAAACAACTTAATGAACTGACAGAAAGAGAA 443
Qy 301 CAAATCCGTCAAGGCTTGAAGAACTCAGAAAGTTCTGCGAGAGAGACACTTAATG 360
Db 444 CAAATCCGTCAAGGCTTGAAGAACTCAGAAAGTTCTGCGAGAGAGACACTTAATG 503
Qy 361 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAAGAGAGTACAGG 420
Db 504 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAAGAGAGTACAGG 563
Qy 421 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTCTAT 480
Db 564 ACAGCCAGGCTCATCTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTCTAT 623
Qy 481 TCAGAGGGAGGCTATAGGCTCGAGATCTTGATGATGTTTCTGTTTGGTGG 540
Db 624 TCAGAGGGAGGCTATAGGCTCGAGATCTTGATGATGTTTCTGTTTGGTGG 683
Qy 541 AAGATTTCAATGAGACAGAGCTGGCCGAGATTGCGAGAGTAAAGATCAATGTTTCCC 600
Db 684 AAGATTTCAATGAGACAGAGCTGGCCGAGATTGCGAGAGTAAAGATCAATGTTTCCC 743
Qy 601 GTGAATGAGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTGTC 660
Db 744 GTGAATGAGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTGTC 803
Qy 661 ATCGAAATTTAGAGAGTGAACCATCAACATATGTCAGAGAGTCAATTTCAAGTTGTC 720
```

```
Db 804 ATGGAATTTCTAGACAGTGAACCATTCATATGTGCAGAGAGATCATTTCAAGTGTGC 863
Qy 721 GTGAGAGAAACGGCTTCCGACATAGCCCGAAGGTGCGCTGCGACCTTCAAG 780
Db 864 GTGAGAGAAACGGCTTCCGACATAGCCCGAAGGTGCGCTGCGACCTTCAAG 923
Qy 781 ATCAATGACTCGGTCACTCAATAGAAACCCCTTTCTGTGAGAGACATTATTTACTG 840
Db 924 ATCAATGACTCGGTCACTCAATAGAAACCCCTTTCTGTGAGAGATATTATTTACTG 983
Qy 841 TGTCCAGCGCTTATCTTAAAGAAAGTTGSCATGAAAGCTGCACTCCAGTCAAGTAAAC 900
Db 984 TGTCCAGCGCTTATCTTAAAGAAAGTTGSCATGAAAGCTGCACTCCAGTCAAGTAAAC 1043
Qy 901 GATGGCCTCTCTTTATCTTATCTCAAGTTCGTATCATATCAACACACATGTTGACGGT 960
Db 1044 GATGGCCTCTCTTTATCTTATCTCAAGTTCGTATCATATCAACACACATGTTGACGGT 1103
Qy 961 TCCATCCTGAGCAATCGCCCTGTGATCTGTCTCTGCTAGCCCTGAGCTCTCTGAG 1020
Db 1104 TCCATCCTGAGCAATCGCCCTGTGATCTGTCTCTGCTAGCCCTGAGCTCTCTGAG 1163
Qy 1021 TGGTTCTGCGCCCTCTGTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCCGAG 1080
Db 1164 TGGTTCTGCGCCCTCTGTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCCGAG 1223
Qy 1081 GAGAGTGAGAGAAA 1094
Db 1224 GAGAGTGAGAGAAA 1237
```

## RESULT 7

```
US-10-474-794-231
; Sequence 231, Application US/10474794
; Publication No. US20040213793A1
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: US/10/474,794
; PRIOR FILING DATE: 2003-10-14
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 231
; LENGTH: 5540
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-474-794-231
```

```
Query Match 98.8%; Score 1090.8; DB 20; Length 5540;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1092; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 ATGGCCACGGCGGAGAGAGCCCTGCGCATCGGCTTCCAGTGCCTCTTTGGCCACT 60
Db 144 ATGGCCACGGCGGAGAGAGCCCTGCGCATCGGCTTCCAGTGCCTCTTTGGCCACT 203
Qy 61 CTGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGGTCCAGCTCTGTAC 120
Db 204 CTGGTCTCATCTGCGCGGCGGAGAGCGGAGAGATGGGGGTCCAGCTCTGTAC 263
Qy 121 GGGCGATTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTGTGACCACTGGAAAT 180
Db 264 GGGCGATTTGACCTGTACTTCAATTTTGGACAAATCAGGAAGTGTGTGACCACTGGAAAT 323
```

```
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAAATTCATACGCCACAGTTGAGATG 240
Db 324 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAAATTCATACGCCACAGTTGAGAAATG 383
Qy 241 TCCCTTATTTGTTTCTCCACCGGAGAAACACCTTAATGAACTGACAGAAACAGAGAA 300
Db 384 TCCCTTATTTGTTTCTCCACCGGAGAAACACCTTAATGAACTGACAGAAACAGAGAA 443
Qy 301 CAAATCCGTCAAGGCTAGAGAAATCCAGAAAGTTGTGCGCAGAGAGACATTAATG 360
Db 444 CAAATCCGTCAAGGCTAGAGAAATCCAGAAAGTTGTGCGCAGAGAGACATTAATG 503
Qy 361 CATGAAAGATTTGAAAGGGCCAGTGACAGATTTATTAAGAAACAGACAGGGTCAAG 420
Db 504 CATGAAAGATTTGAAAGGGCCAGTGACAGATTTATTAAGAAACAGACAGGGTCAAG 563
Qy 421 ACAGCGAGGCTATCATCTGCTTGACTGATGAGAACTCATGAAAGATCTCTTTTCTAT 480
Db 564 ACAGCGAGGCTATCATCTGCTTGACTGATGAGAACTCATGAAAGATCTCTTTTCTAT 623
Qy 481 TCAGAGAGGAGGCTAATAGTCTGAGATCTTGTGCAATTTGTTTACTGTGTGTGTG 540
Db 624 TCAGAGAGGAGGCTAATAGTCTGAGATCTTGTGCAATTTGTTTACTGTGTGTGTG 583
Qy 541 AAAATTTCAATGAGACACAGCTGGCCGGATTTGCGGACAGTAAAGATCATGTGTTCCC 600
Db 684 AAAATTTCAATGAGACACAGCTGGCCGGATTTGCGGACAGTAAAGATCATGTGTTCCC 743
Qy 601 GTGAATGACGGCTTTCAGGCTGCGCAAGGATCATCTCAATTTTGAAGAAAGTCTCTG 660
Db 744 GTGAATGACGGCTTTCAGGCTGCGCAAGGATCATCTCAATTTTGAAGAAAGTCTCTG 803
Qy 661 ATCGAAATTTCTAGCAGCTGAAACCATTCACATATGTGACAGAGAGTCAATTCAGTGTGC 720
Db 804 ATCGAAATTTCTAGCAGCTGAAACCATTCACATATGTGACAGAGAGTCAATTCAGTGTGC 863
Qy 721 GTGAGAGAAACGGCTTCCGACATAGCCCGAAGGTGCGCTGCGACCTTCAAG 780
Db 864 GTGAGAGAAACGGCTTCCGACATAGCCCGAAGGTGCGCTGCGACCTTCAAG 923
Qy 781 ATCAATGACTCGGTCACTCAATAGAAACCCCTTTCTGTGAGAGACATTATTTACTG 840
Db 924 ATCAATGACTCGGTCACTCAATAGAAAGCCCTTTCTGTGAGAGATATTATTTACTG 983
Qy 984 TGTCCAGCGCTTATCTTAAAGAAAGTTGSCATGAAAGCTGCACTCCAGTCAAGTAAAC 1043
Db 1044 TGTCCAGCGCTTATCTTAAAGAAAGTTGSCATGAAAGCTGCACTCCAGTCAAGTAAAC 1103
Qy 1044 GATGGCCTCTCTTTATCTTATCTCAAGTTCGTATCATATCAACACACATGTTGACGGT 1103
Db 961 TCCATCCTGAGCAATCGCCCTGTGATCTGTCTCTGCTAGCCCTGAGCTCTCTGAG 1020
Qy 1104 TCCATCCTGAGCAATCGCCCTGTGATCTGTCTCTGCTAGCCCTGAGCTCTCTGAG 1163
Qy 1021 TGGTTCTGCGCCCTCTGTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCCGAG 1080
Db 1164 TGGTTCTGCGCCCTCTGTGCTGCACTGTGATTATCAAGAGGTCCCTCCACCCCTGCCGAG 1223
Qy 1081 GAGAGTGAGAGAAA 1094
Db 1224 GAGAGTGAGAGAAA 1237
```

## RESULT 8

```
US-10-038-307-17
; Sequence 17, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKANMAK
```

APPLICANT: Judith J. HEALEY  
FILE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038.307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 1674  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-038-307-17

Query Match 87.0%; Score 960.8; DB 15; Length 1674;  
Best Local Similarity 99.8%; Pred. No. 3.9e-303;  
Matches 962; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
Db 13 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 72  
Qy 61 CTGGTGTCTCATCTGCGCGGCGGAGGAGGAGGAGATGGGGGTCCAGCTGTCTAC 120  
Db 73 CTGGTGTCTCATCTGCGCGGCGGAGGAGGAGGAGATGGGGGTCCAGCTGTCTAC 132  
Qy 121 GCGCGATTGACCTGTACTTATTTGACAAATCAGGAAGTGTGTGACCACTGGAAT 180  
Db 133 GCGCGATTGACCTGTACTTATTTGACAAATCAGGAAGTGTGTGACCACTGGAAT 192  
Qy 181 GAAATCTATTACTTTGTGGAACAAGTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 240  
Db 193 GAAATCTATTACTTTGTGGAACAAGTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 252  
Qy 241 TCCCTTATTTGTTTCTCACCCGAGGAACAACCTTAATGAACCTGACAGAGACAGAGAA 300  
Db 253 TCCCTTATTTGTTTCTCACCCGAGGAACAACCTTAATGAACCTGACAGAGACAGAGAA 312  
Qy 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGGACACTTACATG 360  
Db 313 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGGACACTTACATG 372  
Qy 361 CATGAAGATTGTAAGAGGCGGAGTGAAGCAATTTATATGAAAACAGACAGGGTACAG 420  
Db 373 CATGAAGATTGTAAGAGGCGGAGTGAAGCAATTTATATGAAAACAGACAGGGTACAG 432  
Qy 421 ACAGCCAGGCTCATTTGCTTTGACTGATGAGAGAACTCCATGAGAACTCTTTTCTAT 480  
Db 433 ACAGCCAGGCTCATTTGCTTTGACTGATGAGAGAACTCCATGAGAACTCTTTTCTAT 492  
Qy 481 TCAGAGAGGAGGCTAATAGGCTTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 540  
Db 493 TCAGAGAGGAGGCTAATAGGCTTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 552  
Qy 541 AAGGATTCAATGAGACACAGCTGCGCGGATTCGGAAGTGAAGATCATGTGTTTCCC 600  
Db 553 AAGGATTCAATGAGACACAGCTGCGCGGATTCGGAAGTGAAGATCATGTGTTTCCC 612  
Qy 601 GTGAAGTACGGCTTTCAGGCTTCGAGAGGATCATCTCAATTTGAGAAAGTCTGTC 660  
Db 613 GTGAAGTACGGCTTTCAGGCTTCGAGAGGATCATCTCAATTTGAGAAAGTCTGTC 672  
Qy 661 ATGGAATTTCTAGAGCTGAACATTCACCAATATGTGACAGAGAGTCAATTTCAAGTTGTC 720  
Db 673 ATGGAATTTCTAGAGCTGAACATTCACCAATATGTGACAGAGAGTCAATTTCAAGTTGTC 732  
Qy 721 GTGAGAGAAAACGGCTTTCGAGATGCGCGGACAGTGAACAGGGTCTCTGTGAGCTTCAAG 780  
Db 733 GTGAGAGAAAACGGCTTTCGAGATGCGCGGACAGTGAACAGGGTCTCTGTGAGCTTCAAG 792  
Qy 781 ATCAATGACTCGGTCACTCAATGAGAGCCCTTTCTGTGAGAGCACTTAATTTACTG 840  
Db 793 ATCAATGACTCGGTCACTCAATGAGAGCCCTTTCTGTGAGAGCACTTAATTTACTG 852

Qy 841 TGTCCAGCCCTCATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGGTCAAGTAAAC 900  
Db 853 TGTCCAGCCCTCATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGGTCAAGTAAAC 912  
Qy 901 GATGGCTCTCTTTTATCTCAGTTCGTGATCATATCCACCAACACTGTTCTGAAGGT 960  
Db 913 GATGGCTCTCTTTTATCTCAGTTCGTGATCATATCCACCAACACTGTTCTGAAGGT 972  
Qy 961 TCCA 964  
Db 973 CCGA 976

RESULT 9  
US-10-201-292-17  
Sequence 17; Application US/10201292  
Publication No. US2003014193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 1674  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-201-292-17

Query Match 87.0%; Score 960.8; DB 15; Length 1674;  
Best Local Similarity 99.8%; Pred. No. 3.9e-303;  
Matches 962; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
Db 13 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 72  
Qy 61 CTGGTGTCTCATCTGCGCGGCGGAGGAGGAGGAGATGGGGGTCCAGCTGTCTAC 120  
Db 73 CTGGTGTCTCATCTGCGCGGCGGAGGAGGAGGAGATGGGGGTCCAGCTGTCTAC 132  
Qy 121 GCGGATTGACCTGTACTTATTTGACAAATCAGGAAGTGTGTGACCACTGGAAT 180  
Db 133 GCGGATTGACCTGTACTTATTTGACAAATCAGGAAGTGTGTGACCACTGGAAT 192  
Qy 181 GAAATCTATTACTTTGTGGAACAAGTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 240  
Db 193 GAAATCTATTACTTTGTGGAACAAGTGGCTCAAAATTCATCAGCCCAAGTTGAAGATG 252  
Qy 241 TCCCTTATTTGTTTCTCACCCGAGGAACAACCTTAATGAACCTGACAGAGACAGAGAA 300  
Db 253 TCCCTTATTTGTTTCTCACCCGAGGAACAACCTTAATGAACCTGACAGAGACAGAGAA 312  
Qy 301 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGGACACTTACATG 360  
Db 313 CAAATCCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGGACACTTACATG 372  
Qy 361 CATGAAGATTGTAAGAGGCGGAGTGAAGCAATTTATATGAAAACAGACAGGGTACAG 420  
Db 373 CATGAAGATTGTAAGAGGCGGAGTGAAGCAATTTATATGAAAACAGACAGGGTACAG 432  
Qy 421 ACAGCCAGGCTCATTTGCTTTGACTGATGAGAGAACTCCATGAGAAATCTTTTCTAT 480  
Db 433 ACAGCCAGGCTCATTTGCTTTGACTGATGAGAGAACTCCATGAGAAATCTTTTCTAT 492  
Qy 481 TCAGAGAGGAGGCTAATAGGCTTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 540  
Db 493 TCAGAGAGGAGGCTAATAGGCTTCGAGATCTTGGTCAATTTGTTACTGTGTGTG 552

Qy 541 AAAGATTTCATGAGACAGAGCTGGCCCGGATTCGCGACAGTAAGATCATGTGTTTCCC 600  
Db 553 AAAGATTTCATGAGACAGAGCTGGCCCGGATTCGCGACAGTAAGATCATGTGTTTCCC 612  
Qy 601 GTGAAATGACGGCTTTGAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAAGTCTGC 660  
Db 613 GTGAAATGACGGCTTTGAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAAGTCTGC 672  
Qy 661 ATGGAATTTGAGAGCTGAACCATCCACATATGTCAGAGAGATGATTTCAAGTTGTC 720  
Db 673 ATGGAATTTGAGAGCTGAACCATCCACATATGTCAGAGAGATGATTTCAAGTTGTC 732  
Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGCAAGTTCAAG 780  
Db 733 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGCAAGTTCAAG 792  
Qy 781 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTTCTGTGGAAGACATTAATTA 840  
Db 793 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTTCTGTGGAAGATTAATTA 852  
Qy 841 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAGCTGACCTCCAGTCAGCATGAC 900  
Db 853 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAGCTGACCTCCAGTCAGCATGAC 912  
Qy 901 GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACACTGTTCTGACG 960  
Db 913 GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACTGTTCTGACG 972  
Qy 961 TCCTA 964  
Db 973 CCCA 976

RESULT 10  
US-10-038-307-13  
; Sequence 13, Application US/10038307  
; Publication No. US20030134786A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-038-307-13

Query Match 86.3%, Score 952.4, DB 15, Length 1650;  
Best Local Similarity 99.4%, Pred. No. 2.2e-300;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
Db 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
Qy 61 CTGGGCTCATCTGGCGCGGAGAGAGGAGGAGATGGGGGTCCAGCTGCTAC 120  
Db 61 CTGGGCTCATCTGGCGCGGAGAGAGGAGGAGATGGGGGTCCAGCTGCTAC 120  
Qy 121 GGGCGAATTTGACTGTACTTATTTTGAACAATCAGAAAGTGTGTGCACTGGAAT 180  
Db 121 GGGCGAATTTGACTGTACTTATTTTGAACAATCAGAAAGTGTGTGCACTGGAAT 180  
Qy 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAAATTCATCAGCCCAAGTTGAAGATG 240  
Db 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAAATTCATCAGCCCAAGTTGAAGATG 240

Db 181 GAAATCTATTACTTTGTGGAACAGTTGGCTCACAAAATTCATCAGCCCAAGTTGAAGATG 240  
Qy 241 TCCTTATTTGTTTCTCCACCCGAGAACACTTAATGAACTGA CAGAGACAGAGAA 300  
Db 241 TCCTTATTTGTTTCTCCACCCGAGAACACTTAATGAACTGA CAGAGACAGAGAA 300  
Qy 301 CAAATCCGTCAAGGCTTAGAAGATCTCCAGAAAGTTCTGCGAGAGAGACATTAATG 360  
Db 301 CAAATCCGTCAAGGCTTAGAAGATCTCCAGAAAGTTCTGCGAGAGAGACATTAATG 360  
Qy 361 CATTGAAGATTTGAAGAGGCGAGTGAAGATTTATATGAAAAAGAGAGGATCAGG 420  
Db 361 CATTGAAGATTTGAAGAGGCGAGTGAAGATTTATATGAAAAAGAGAGGATCAGG 420  
Qy 421 ACAGCCAGGCTCATCTTGTGACTGATGAGAACTCCATGAAAGATCTTTTCTAT 480  
Db 421 ACAGCCAGGCTCATCTTGTGACTGATGAGAACTCCATGAAAGATCTTTTCTAT 480  
Qy 481 TCAGAGAGGAGGCTTAATGAGTCTGAGATCTTTGTGCAATTTGTATGTGTG 540  
Db 481 TCAGAGAGGAGGCTTAATGAGTCTGAGATCTTTGTGCAATTTGTATGTGTG 540  
Qy 541 AAAGATTTCATGAGACAGAGCTGGCCCGGATTCGCGACAGTAAGATCATGTGTTTCCC 600  
Db 541 AAAGATTTCATGAGACAGAGCTGGCCCGGATTCGCGACAGTAAGATCATGTGTTTCCC 600  
Qy 601 GTGAAATGACGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAAAGTCTGC 660  
Db 601 GTGAAATGACGGCTTTGAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAAAGTCTGC 660  
Qy 661 ATGGAATTTGAGAGCTGAACCATCACTCAATATGTCAGAGAGATTTTCAAGTTGTC 720  
Db 661 ATGGAATTTGAGAGCTGAACCATCACTCAATATGTCAGAGAGATTTTCAAGTTGTC 720  
Qy 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGCAAGTTCAAG 780  
Db 721 GTGAGAGAAACGGCTTCCGACATGCGCGCAACGTGACAGGGTCTCTGCAAGTTCAAG 780  
Qy 781 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTTCTGTGGAAGACATTAATTA 840  
Db 781 ATCAATGACTCGGTCACTCACTCAATGAGAAAGCCCTTTTCTGTGGAAGATTAATTA 840  
Qy 841 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAGCTGCACTCCAGTCAGCATGAC 900  
Db 841 TGTCCAGCGCTATCTTAAAGAAAGTTGGCATGAAGCTGCACTCCAGTCAGCATGAC 900  
Qy 901 GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACTGTTCTGACG 960  
Db 901 GATGGCTCTCTTTTATCTCCAGTTCTGTCAATCAACACACACTGTTCTGACG 960  
Qy 961 TC 962  
Db 961 TC 962

RESULT 11  
US-10-038-307-15  
; Sequence 15, Application US/10038307  
; Publication No. US20030134786A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/038,307  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 1650  
; TYPE: DNA

! ORGANISM: Homo sapiens  
US-10-038-307-15

Query Match 86.3%; Score 952.4; DB 15; Length 1650;  
Best Local Similarity 99.4%; Pred. No. 2.2e-300;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60
DB 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60
QY 61 CTGGGCTCATCTGGCCGGGCAAGGGGAGCGCAGGAGATGGGGGCTCCAGCTGCTAC 120
DB 61 CTGGGCTCATCTGGCCGGGCAAGGGGAGCGCAGGAGATGGGGGCTCCAGCTGCTAC 120
QY 121 GCGGATTTGACCTGTACTTCTTTGAGCAAAATCAGGAAGTGTCTGCACCACTGGAAT 180
DB 121 GCGGATTTGACCTGTACTTCTTTGAGCAAAATCAGGAAGTGTCTGCACCACTGGAAT 180
QY 181 GAAATCTAATCTTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAATG 240
DB 181 GAAATCTAATCTTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAATG 240
QY 241 TCCCTTAATGTTTTCTCACCAGGAAACAACCTTAATGAACCTGACGAAGACAGAGAA 300
DB 241 TCCCTTAATGTTTTCTCACCAGGAAACAACCTTAATGAACCTGACGAAGACAGAGAA 300
QY 301 CAAATCCGTCAGAGGCTAGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360
DB 301 CAAATCCGTCAGAGGCTAGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360
QY 361 CATGAAGATTTGAAGGGCCAGTGAAGCAATTAATGAAGAAACAGCAAGGGTACAG 420
DB 361 CATGAAGATTTGAAGGGCCAGTGAAGCAATTAATGAAGAAACAGCAAGGGTACAG 420
QY 421 ACAGCCAGGTCATCATTTGCTTTGACTGATGAGAACTTCCATGAAGATCTCTTTTCTAT 480
DB 421 ACAGCCAGGTCATCATTTGCTTTGACTGATGAGAACTTCCATGAAGATCTCTTTTCTAT 480
QY 481 TCAGAGAGGAGGCTAATAGGCTCGAGATCTGGTGCAATTTGTTTACTGTGTTGATG 540
DB 481 TCAGAGAGGAGGCTAATAGGCTCGAGATCTGGTGCAATTTGTTTACTGTGTTGATG 540
QY 541 AAAGATTTCAATGAGACACAGCTGGCCCGGATGGCGGACAGTAAGATCATGTGTTCCC 600
DB 541 AAAGATTTCAATGAGACACAGCTGGCCCGGATGGCGGACAGTAAGATCATGTGTTCCC 600
QY 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCACTCAATTTTGAAGAGTCTGCG 660
DB 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCACTCAATTTTGAAGAGTCTGCG 660
QY 661 ATCGAAATTTCTAGCAGCTGAACATCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720
DB 661 ATCGAAATTTCTAGCAGCTGAACATCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720
QY 721 GTGAGAGGAAACGGCTTCCGACATGCCCCGCAACGTGACAGGCTCTTGCAGCTTCAAG 780
DB 721 GTGAGAGGAAACGGCTTCCGACATGCCCCGCAACGTGACAGGCTCTTGCAGCTTCAAG 780
QY 781 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTTCTGTGAAAGACATTAATTTACTG 840
DB 781 ATCAATGACTCGGTCACTCAATGAGAAAGCCCTTTTCTGTGAAAGACATTAATTTACTG 840
QY 841 TGTCCAGGCGCTATCTTAAAGAAAGTTGGCATGAAAGTGAAGTCACTCCAGGTACAGTAAC 900
DB 841 TGTCCAGGCGCTATCTTAAAGAAAGTTGGCATGAAAGTGAAGTCACTCCAGGTACAGTAAC 900
QY 901 GATGGCCTCTTTTATCTCCAGTTCTGTCAATCAACCACTGTTCTGACGGT 960
DB 901 GATGGCCTCTTTTATCTCCAGTTCTGTCAATCAACCACTGTTCTGACGGT 960
QY 961 TC 962
DB 961 TC 962
```

DB 961 TC 962

RESULT 12  
US-10-201-292-13  
; Sequence 13, Application US/10201292  
; Publication No. US20030144193A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTHMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKANMAK  
; APPLICANT: Judith J. HENLEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/201,292  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-201-292-13

Query Match 86.3%; Score 952.4; DB 15; Length 1650;  
Best Local Similarity 99.4%; Pred. No. 2.2e-300;  
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60
DB 1 ATGGCCACGGCGGAGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60
QY 61 CTGGGCTCATCTGGCCGGGCAAGGGGAGCGCAGGAGATGGGGGCTCCAGCTGCTAC 120
DB 61 CTGGGCTCATCTGGCCGGGCAAGGGGAGCGCAGGAGATGGGGGCTCCAGCTGCTAC 120
QY 121 GCGGATTTGACCTGTACTTCTTTGAGCAAAATCAGGAAGTGTCTGCACCACTGGAAT 180
DB 121 GCGGATTTGACCTGTACTTCTTTGAGCAAAATCAGGAAGTGTCTGCACCACTGGAAT 180
QY 181 GAAATCTAATCTTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAATG 240
DB 181 GAAATCTAATCTTTGTGGAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAAATG 240
QY 241 TCCCTTAATGTTTTCTCACCAGGAAACAACCTTAATGAACCTGACGAAGACAGAGAA 300
DB 241 TCCCTTAATGTTTTCTCACCAGGAAACAACCTTAATGAACCTGACGAAGACAGAGAA 300
QY 301 CAAATCCGTCAGAGGCTAGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360
DB 301 CAAATCCGTCAGAGGCTAGAAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360
QY 361 CATGAAGATTTGAAGGGCCAGTGAAGCAATTAATGAAGAAACAGCAAGGGTACAG 420
DB 361 CATGAAGATTTGAAGGGCCAGTGAAGCAATTAATGAAGAAACAGCAAGGGTACAG 420
QY 421 ACAGCCAGGTCATCATTTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTTCTAT 480
DB 421 ACAGCCAGGTCATCATTTGCTTTGACTGATGAGAACTCCATGAAGATCTCTTTTCTAT 480
QY 481 TCAGAGAGGAGGCTAATAGGCTCGAGATCTTGTGCAATTTGTTTACTGTGTTGATG 540
DB 481 TCAGAGAGGAGGCTAATAGGCTCGAGATCTTGTGCAATTTGTTTACTGTGTTGATG 540
QY 541 AAAGATTTCAATGAGACACAGCTGGCCCGGATGGCGGACAGTAAGATCATGTGTTCCC 600
DB 541 AAAGATTTCAATGAGACACAGCTGGCCCGGATGGCGGACAGTAAGATCATGTGTTCCC 600
QY 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCACTCAATTTTGAAGAGTCTGCG 660
DB 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGATCATCTCACTCAATTTTGAAGAGTCTGCG 660
QY 661 ATCGAAATTTCTAGCAGCTGAACATCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720
DB 661 ATCGAAATTTCTAGCAGCTGAACATCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720
```



```
Db 661 ATGGAATTCTAGAGAGCTGAACCATCACTATGTGACAGAGAGTCAATTCAGTGTGTC
Qy 721 GTGAGAGAAACGGCTTCGCGCATGCCCCGAACGCTGACAGGGCTCTGACAGCTTCAAG
Db 721 GTGAGAGAAACGGCTTCGCGCATGCCCCGAACGCTGACAGGGCTCTGACAGCTTCAAG
Qy 781 ATCAATGACTGCTGACACTCACTCAATGAGAGAGCCCTTTCTGTGAGAACACTTATTACTG
Db 781 ATCAATGACTGCTGACACTCACTCAATGAGAGAGCCCTTTCTGTGAGAACACTTATTACTG
Qy 841 TGTCACAGGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTCAAGTCAATGAC
Db 841 TGTCACAGGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTCAAGTCAATGAC
Qy 901 GATGGCCCTCTTTTATCTTATCTGAGTTCTGTGATCATCAACCAACACACTGTCTGACGGT
Db 901 GATGGCCCTCTTTTATCTTATCTGAGTTCTGTGATCATCAACCAACACACTGTCTCCAAA
Qy 961 TC 962
Db 961 TC 962
```

```
RESULT 13
US-10-201-292-15
; Sequence 15, Application US/10201292
; Publication No. US20030144193A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/201,292
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-201-292-15
```

```
Query Match 86.3%; Score 952.4; DB 15; Length 1650;
Best Local Similarity 99.4%; Pred. No. 2.2e-300;
Matches 956; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATTCGGCTTCCAGTGGCTCTCTTTGGCCACT
Db 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATTCGGCTTCCAGTGGCTCTCTTTGGCCACT
Qy 61 CTGGTGCATCTGTGCGCGGGGAGAGGAGGAGGAGATGGGGGCTTCAGCTGTCTAC 120
Db 61 CTGGTGCATCTGTGCGCGGGGAGAGGAGGAGGAGATGGGGGCTTCAGCTGTCTAC 120
Qy 121 GCGCGATTGACCTGTACTCTTCACTTTTGGCAAAATCAGAGAGTGTGCTCACCCTGGAAT
Db 121 GCGCGATTGACCTGTACTCTTCACTTTTGGCAAAATCAGAGAGTGTGCTCACCCTGGAAT
Qy 181 GAAATCTATTACTTTGTGGAACAAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG
Db 181 GAAATCTATTACTTTGTGGAACAAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG
Qy 241 TCCTTATTTGTTTTCTCCACCCGAGAGAACCTTAATGAAATGACGAAAGACAGAGAA
Db 241 TCCTTATTTGTTTTCTCCACCCGAGAGAACCTTAATGAAATGACGAAAGACAGAGAA
Qy 301 CAAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGAGACACTTAACATG
Db 301 CAAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCGAGGAGAGACACTTAACATG
```

```
Qy 361 CATGAAGATTGTAAGAGGCGCACTGAGCAAGATTATTATGAAAAAGACAGAGGTTACAG
Db 361 CATGAAGATTGTAAGAGGCGCGAGTGAGAGATTATTATGAAAAAGACAGAGGTTACAG
Qy 421 ACAGCAGGCTCATCTATGCTTTGACCTGATGAGAACTCCATGAAATGCTCTTTCTAT
Db 421 ACAGCAGGCTCATCTATGCTTTGACCTGATGAGAACTCCATGAAATGCTCTTTCTAT
Qy 481 TCAGAGAGGAGGCTAATAGTCTCGAGATCTTGTGTGCAATGTTTACTGTGTGTG
Db 481 TCAGAGAGGAGGCTAATAGTCTCGAGATCTTGTGTGCAATGTTTACTGTGTGTG
Qy 541 AAGATTTCATGAGACACAGCTGCCCCGAGATTGGCGACATGAGATCATGTGTTCCC
Db 541 AAGATTTCATGAGACACAGCTGCCCCGAGATTGGCGACATGAGATCATGTGTTCCC
Qy 601 GTAATGACGGCTTTAGGCTGCAAGGATCATCACTCAATTTTGAAGAGTCCGTC
Db 601 GTAATGACGGCTTTAGGCTGCAAGGATCATCACTCAATTTTGAAGAGTCCGTC
Qy 661 ATGGAATTCTAGACAGCTGAACCATCCACATATGTGACAGAGATCATTTCAAGTTGTC
Db 661 ATGGAATTCTAGACAGCTGAACCATCCACATATGTGACAGAGATCATTTCAAGTTGTC
Qy 721 GTGAGAGAAACGGCTTCGCACTGCCCGCAAGTGAACAGGCTCTCTGACCTTCAAG
Db 721 GTGAGAGAAACGGCTTCGCACTGCCCGCAAGTGAACAGGCTCTCTGACCTTCAAG
Qy 781 ATCAATGACTGCTGACACTCACTCAATGAGAGAGCCCTTTCTGTGAGAACACTTATTACTG
Db 781 ATCAATGACTGCTGACACTCACTCAATGAGAGAGCCCTTTCTGTGAGAACACTTATTACTG
Qy 841 TGTCACAGGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTCAAGTCAATGAC
Db 841 TGTCACAGGCTTATCTTAAAGAAAGTTGGCATGAAAGCTGCACTCCAGTCAAGTCAATGAC
Qy 901 GATGGCCCTCTTTTATCTTATCTGACGTTCTGTATCATCAACCAACACACTGTCTGACGGT
Db 901 GATGGCCCTCTTTTATCTTATCTGACGTTCTGTATCATCAACCAACACACTGTCTCCAAA
Qy 961 TC 962
Db 961 TC 962
```

```
RESULT 14
US-10-038-307-9
; Sequence 9, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-038-307-9
```

```
Query Match 86.1%; Score 950.8; DB 15; Length 1650;
Best Local Similarity 99.3%; Pred. No. 7.3e-300;
Matches 955; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATTCGGCTTCCAGTGGCTCTCTTTGGCCACT
Db 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATTCGGCTTCCAGTGGCTCTCTTTGGCCACT
```



QY 61 CTGGTCTCATCTGCGCGGCGAAGGCGGAGGAGATGGGGGCTCCAGCTGCTAC 120  
 DB 61 CTGGTCTCATCTGCGCGGCGAAGGCGGAGGAGATGGGGGCTCCAGCTGCTAC 120  
 QY 121 GCGGATTTGACCTGTACTCTCTTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180  
 DB 121 GCGGATTTGACCTGTACTCTCTTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180  
 QY 181 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCCACTGTAAGATG 240  
 DB 181 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCCACTGTAAGATG 240  
 QY 241 TCCCTTATTGTTTCTCCACCCGAGAACCACTTAAATGAAACTGACAGAAAGACAGAA 300  
 DB 241 TCCCTTATTGTTTCTCCACCCGAGAACCACTTAAATGAAACTGACAGAAAGACAGAA 300  
 QY 301 CAATTCCTTCAGAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360  
 DB 301 CAATTCCTTCAGAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360  
 QY 361 CATGAAGATTGAAAGGCGCAGTAGACAGATTATTAATGAAAAGACAGAGGATACAG 420  
 DB 361 CATGAAGATTGAAAGGCGCAGTAGACAGATTATTAATGAAAAGACAGAGGATACAG 420  
 QY 421 ACAGCCAGCGTCATCATCTGCTTTGACAGTAGAGAACTCCAGAAAGATCTCTTTTCTAT 480  
 DB 421 ACAGCCAGCGTCATCATCTGCTTTGACAGTAGAGAACTCCAGAAAGATCTCTTTTCTAT 480  
 QY 481 TCAGAGAGGAGGCTTAATAGGCTCTGAGATCTTGGTCAATTTGCTGTGTGTG 540  
 DB 481 TCAGAGAGGAGGCTTAATAGGCTCTGAGATCTTGGTCAATTTGCTGTGTGTG 540  
 QY 541 AAAAGATTCAATGAGACACAGCTGGCCGGATTTGGGAGACATGATGATGTTTCC 600  
 DB 541 AAAAGATTCAATGAGACACAGCTGGCCGGATTTGGGAGACATGATGATGTTTCC 600  
 QY 601 GTGAATGACGGCTTCAAGCTCTGCAAGGACATCATCACTCAATTTTGAAGAGTCTGCG 660  
 DB 601 GTGAATGACGGCTTCAAGCTCTGCAAGGACATCATCACTCAATTTTGAAGAGTCTGCG 660  
 QY 661 ATCGAAATTTCTAGCAGCTGAAACATCCACATATGTGAGAGAGTCACTTCAAGTGTG 720  
 DB 661 ATCGAAATTTCTAGCAGCTGAAACATCCACATATGTGAGAGAGTCACTTCAAGTGTG 720  
 QY 721 GTGAGAGGAAACGGCTTCGACATGCGCGAACGTCGACAGGCTCTCTGACGCTTCAAG 780  
 DB 721 GTGAGAGGAAACGGCTTCGACATGCGCGAACGTCGACAGGCTCTCTGACGCTTCAAG 780  
 QY 781 ATCAATGACCTGCTGACACTCAATGAGAGCCCTTTCTGTGAGAGATCTTATTTACTG 840  
 DB 781 ATCAATGACCTGCTGACACTCAATGAGAGCCCTTTCTGTGAGAGATCTTATTTACTG 840  
 QY 841 TGTCCAGCGCTATTTTAAAGAGTGGCATGAAAGCTGCACTCCAGSTCAGCTGAAC 900  
 DB 841 TGTCCAGCGCTATTTTAAAGAGTGGCATGAAAGCTGCACTCCAGSTCAGCTGAAC 900  
 QY 901 GATGCGCTCTCTTATCTCAGATTCTGATCAATCAACCAACCACTCTTCTCCAA 960  
 DB 901 GATGCGCTCTCTTATCTCAGATTCTGATCAATCAACCAACCACTCTTCTCCAA 960  
 QY 961 TC 962  
 DB 961 TC 962

RESULT 15  
 US-10-201-292-9  
 ; Sequence 9, Application US/10201292  
 ; Publication No. US20030144193A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: James B. ROTHMAN  
 ; APPLICANT: Theresa L. O'KEEFE

; APPLICANT: Engin OZKANAK  
 ; APPLICANT: Judith J. HEALEY  
 ; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
 ; FILE REFERENCE: 7853-253-999  
 ; CURRENT APPLICATION NUMBER: US/10/201,292  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 1650  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-201-292-9  
 Query Match 86.1%; Score 950.8; DB 15; Length 1650;  
 Best Local Similarity 99.3%; Pred. No. 7,3e-300;  
 Matches 955; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
 QY 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
 DB 1 ATGGCCACGGCGGAGCGGAGAGCCCTCGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
 QY 61 CTGGTCTCATCTGCGCGGCGAAGGCGGAGGAGATGGGGGCTCCAGCTGCTAC 120  
 DB 61 CTGGTCTCATCTGCGCGGCGAAGGCGGAGGAGATGGGGGCTCCAGCTGCTAC 120  
 QY 121 GCGGATTTGACCTGTACTCTCTTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180  
 DB 121 GCGGATTTGACCTGTACTCTCTTTTGGACAAATCAGGAAGTGTCTGACCACTGGAAT 180  
 QY 181 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCCACTGTAAGATG 240  
 DB 181 GAAATCTATTACTTGTGGACAGTTGGCTCAAAATTCATCAGGCCCACTGTAAGATG 240  
 QY 241 TCCCTTATTGTTTCTCCACCCGAGAACCACTTAAATGAAACTGACAGAAAGACAGAA 300  
 DB 241 TCCCTTATTGTTTCTCCACCCGAGAACCACTTAAATGAAACTGACAGAAAGACAGAA 300  
 QY 301 CAATTCCTTCAGAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360  
 DB 301 CAATTCCTTCAGAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGACACTTACATG 360  
 QY 361 CATGAAGATTGAAAGGCGCAGTAGACAGATTATTAATGAAAAGACAGAGGATACAG 420  
 DB 361 CATGAAGATTGAAAGGCGCAGTAGACAGATTATTAATGAAAAGACAGAGGATACAG 420  
 QY 421 ACAGCCAGCGTCATCATCTGCTTTGACAGTAGAGAACTCCAGAAAGATCTCTTTTCTAT 480  
 DB 421 ACAGCCAGCGTCATCATCTGCTTTGACAGTAGAGAACTCCAGAAAGATCTCTTTTCTAT 480  
 QY 481 TCAGAGAGGAGGCTTAATAGGCTCTGAGATCTTGGTCAATTTGCTGTGTGTG 540  
 DB 481 TCAGAGAGGAGGCTTAATAGGCTCTGAGATCTTGGTCAATTTGCTGTGTGTG 540  
 QY 541 AAAAGATTCAATGAGACACAGCTGGCCGGATTTGGGAGACATGATGATGTTTCC 600  
 DB 541 AAAAGATTCAATGAGACACAGCTGGCCGGATTTGGGAGACATGATGATGTTTCC 600  
 QY 601 GTGAATGACGGCTTCAAGCTCTGCAAGGACATCATCACTCAATTTTGAAGAGTCTGCG 660  
 DB 601 GTGAATGACGGCTTCAAGCTCTGCAAGGACATCATCACTCAATTTTGAAGAGTCTGCG 660  
 QY 661 ATCGAAATTTCTAGAGCTGAAACATCCACATATGTGAGAGAGTCACTTCAAGTGTG 720  
 DB 661 ATCGAAATTTCTAGAGCTGAAACATCCACATATGTGAGAGAGTCACTTCAAGTGTG 720  
 QY 721 GTGAGAGGAAACGGCTTCGACATGCGCGAACGTCGACAGGCTCTCTGACGCTTCAAG 780  
 DB 721 GTGAGAGGAAACGGCTTCGACATGCGCGAACGTCGACAGGCTCTCTGACGCTTCAAG 780  
 QY 781 ATCAATGACCTGCTGACACTCAATGAGAGCCCTTTCTGTGAGAGACATTTATTTACTG 840  
 DB 781 ATCAATGACCTGCTGACACTCAATGAGAGCCCTTTCTGTGAGAGATCTTATTTACTG 840

Qy	841	TGTCAGGCGCTTATCTTAAAGAAGTTGGCATGAAGCTGCACTCCAGTCAGCATGAAC	900
Db	841	TGTCAGGCGCTTATCTTAAAGAAGTTGGCATGAAGCTGCACTCCAGTCAGCATGAAC	900
Qy	901	GATGGCCTCTCTTTTATCTCCAGTTCTGTCAATCAACCAACACACTGTTCTGACGGT	960
Db	901	GATGGCCTCTCTTTTATCTCCAGTTCTGTCAATCAACCAACACACTGTTCTGACGGT	960
Qy	961	TC 962	
Db	961	TC 962	

Search completed: June 14, 2005, 18:37:32  
 Job time : 786 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2005, 13:05:23 ; Search time 4251 Seconds  
(without alignments)  
9885.428 Million cell updates/sec

Title: US-09-970-076-1\_COPY\_104\_1207

Perfect score: 1104  
Sequence: 1 atggcaccgagcgagcgag.....gtgaggaataataaaaa 1104

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_g881: \*  
9: gb\_g882: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	949.4	86.0	1957	3	CR601167 full-length
2	853.4	77.3	3161	3	AK031465 Mus muscu
3	689.2	62.0	1006	1	AL542724 AL542724
4	684.6	62.0	878	4	BI823853 603039031
5	681.4	61.7	1614	3	AK013005 Mus muscu
6	662	60.0	1054	4	BM800898 AGENCOURT
7	658.8	59.7	833	7	CO245219 AGENCOURT
8	637.8	57.8	963	2	BE741333 601594179
9	585	53.0	725	4	BG326444 602425082
10	582.6	52.8	820	4	BG281561 602402412
11	552	50.0	641	7	CF132432 UI-HF-F00
12	541.4	49.0	780	7	CF126791 UI-HF-ET0
13	534	48.4	823	5	BU131302 603117191
14	533.4	48.3	564	6	CD664532 92B4 IMR3
15	531.4	48.1	877	5	CD664532 92B4 IMR3
16	526.2	47.7	682	7	BX364961 BX364961
17	520.2	47.1	727	6	CA324840 UI-M-F00-
18	507.4	46.0	522	6	CD664533 91G4 IMR3
19	498.4	45.1	600	6	CD664531 91H7 IMR3
20	482	43.7	964	4	BG281831 602403057
21	467.4	42.3	3178	3	AK048637 Mus muscu
22	466.6	42.3	473	7	CY028654 7224 Full
23	464.6	42.1	642	7	CN461036 UI-M-HB0-
24	461	41.8	600	7	CN262102 170005326

C	25	451.6	40.9	643	5	BU632284 UI-H-FE1-
	26	447.4	40.5	638	2	BE369415 601220816
	27	446.8	40.5	582	2	BK470942 DKF2P6861
	28	441.8	40.0	805	6	CB990959 AGENCOURT
	29	437.4	38.7	793	5	BX369960 BX369960
	30	421.2	38.2	867	5	BX337000 BX337000
	31	420.4	38.1	513	2	BE146084 MRO-HT020
	32	420.4	37.7	602	5	BP456257 BP456257
	33	416.4	37.1	524	6	CA945449 UI-M-F00-
	34	409.6	37.0	775	7	CK787494 AGENCOURT
	35	408.2	36.2	465	2	BE145958 MRO-HT020
	36	399.8	34.9	487	2	BE146355 MRO-HT020
	37	382	34.6	775	5	BU363718 603788453
	38	381.2	34.5	539	1	A1466645 mg81d04.Y
	39	376.4	34.1	693	4	BG323538 602422195
	40	373.6	33.8	554	1	A1084806 0280d12.X
C	41	373.6	32.1	546	6	CB519323 UI-M-GH0-
	42	354.4	30.7	592	7	CF728459 UI-M-HB0-
	43	348.6	30.3	736	6	CB724689 EST0528 R
	44	334.2	29.9	356	2	BE146135 MRO-HT020
	45	329.8	29.9	356	2	BE146135 MRO-HT020

#### ALIGNMENTS

RESULT 1  
LOCUS CR601167  
DEFINITION full-length cDNA clone CSOD1035VAL3 of Placenta Cot 25-normalized  
of Homo sapiens (human).  
ACCESSION CR601167  
VERSION CR601167.1 GI:50481974  
KEYWORDS HTCC; CNS/LT; CDNA  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Pull-length cDNA libraries and normalization  
JOURNAL Unpublished  
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/InvitrogenCorporation1600  
Faraday Avenue  
2 (bases 1 to 1957)

REFERENCE  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT  
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime  
end enriched, double-strand cDNA was digested with Not I and cloned  
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library  
was normalized. Library was constructed by Life Technologies, a  
division of Invitrogen.

FEATURES  
source location/Qualifiers  
1..1957  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSOD1035VAL3"  
/issue\_type="Placenta Cot 25-normalized"  
/plasmid="pCMVSPORT\_6"

ORIGIN  
Query Match 86.0%; Score 949.4; DB 3; Length 1957;  
Best Local Similarity 99.9%; Pred. No. 2.1e-270;  
Matches 950; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCCACGGCGGAGAGCCCTTGGCATGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
DB 195 ATGGCCACGGCGGAGAGCCCTTGGCATGGCTTCCAGTGGCTCTCTTTGGCCACT 254

QY 61 CTGGTGTCTCATCTGCGCGGCAAGGGGAGCGAGGAGATGGGGGCTCCAGCTGTCTAC 120  
DB 255 CTGGTGTCTCATCTGCGCGGCAAGGGGAGCGAGGAGATGGGGGCTCCAGCTGTCTAC 314  
QY 121 GGGCGATTTGACCTGTACTTCTTATTTGGAACAATCAGAAAGTGTGTCCACCTGGAAT 180  
DB 315 GGGCGATTTGACCTGTACTTCTTATTTGGAACAATCAGAAAGTGTGTCCACCTGGAAT 374  
QY 181 GAATCTATTACTTGTGGGACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 240  
DB 375 GAATCTATTACTTGTGGGACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 434  
QY 241 TCTTTTATTTGTTTTCTCCACCCGAGGAACAACCTTAATGAACCTGACAGAGAGAGAA 300  
DB 435 TCTTTTATTTGTTTTCTCCACCCGAGGAACAACCTTAATGAACCTGACAGAGAGAGAA 494  
QY 301 CAAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGAGACATTTACATG 360  
DB 495 CAAATTCGTCAAGGCTTGAAGAACTCCAGAAAGTTCTGCCAGAGAGAGACATTTACATG 554  
QY 361 CATGAAGATTTGAAGGGCCAGAGAGAGATTTATTAAGAAACAGACAAAGGGTTACAGG 420  
DB 555 CATGAAGATTTGAAGGGCCAGAGAGAGATTTATTAAGAAACAGACAAAGGGTTACAGG 614  
QY 421 ACAGCCAGGCTCATCATTTGCTTGTGATGATGAGAGAACTCCATGAAGATCTCTTTTCTAT 480  
DB 615 ACAGCCAGGCTCATCATTTGCTTGTGATGATGAGAGAACTCCATGAAGATCTCTTTTCTAT 674  
QY 481 TCAGAGAGGAGGCTTAATAGCTCTCGAGATCTTGGTGCATTTGTTACTGTGTGGTG 540  
DB 675 TCAGAGAGGAGGCTTAATAGCTCTCGAGATCTTGGTGCATTTGTTACTGTGTGGTG 734  
QY 541 AAAGATTTCAATGAGACAGCTGGCCGGATGGCGAGATGGCGACATGAAGATCATGTGTTCCC 600  
DB 735 AAAGATTTCAATGAGACAGCTGGCCGGATGGCGAGATGGCGACATGAAGATCATGTGTTCCC 794  
QY 601 GTGAATGACGGCTTTCAGGCTCTGCAAGGCAATCACTCAATTTTGAAGAGTCTCTGC 660  
DB 795 GTGAATGACGGCTTTCAGGCTCTGCAAGGCAATCACTCAATTTTGAAGAGTCTCTGC 854  
QY 661 ATCGAATTTCTAGACAGCTGAACCTCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 720  
DB 855 ATCGAATTTCTAGACAGCTGAACCTCCACATATGTGAGAGAGTCAATTTCAAGTTGTC 914  
QY 721 GTGAGAGGAAACGGCTTCCGACATGCGCGCAACGCGAGGTCCTTGCAGCTTCAAG 780  
DB 915 GTGAGAGGAAACGGCTTCCGACATGCGCGCAACGCGAGGTCCTTGCAGCTTCAAG 974  
QY 781 ATCAATGACCTGCTCACACTCAATGAGAAAGCCCTTTTCTGTGAGAGACATTTATTTACTG 840  
DB 975 ATCAATGACCTGCTCACACTCAATGAGAAAGCCCTTTTCTGTGAGAGATTTATTTACTG 1034  
QY 841 TGTCCAGGCGCTATTTTAAAGAAAGTTGGCATGAAGCTGCATCTCCAGTCAAGATGAAC 900  
DB 1035 TGTCCAGGCGCTATTTTAAAGAAAGTTGGCATGAAGCTGCATCTCCAGTCAAGATGAAC 1094  
QY 901 GATGGCCTCTGCTTATCTCCAGTTCTGTGCATCATCACACACACACTGT 951  
DB 1095 GATGGCCTCTGCTTATCTCCAGTTCTGTGCATCATCACACACACACTGT 1145

## RESULT 2

AK031465

LOCUS AK031465 3161 bp mRNA linear HTC 03-APR-2004

DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length  
enriched library, clone:6030436p19 product:ANTPAX TOXIN RECEPTOR  
PRECURSOR (TUMOR ENDOTHELIAL MARKER 8) homolog [Mus musculus], full  
insert sequence.

## ACCESSION

AK031465 AK031465.1 GI:26082340

## VERSION

AK031465

## KEYWORDS

HTC; CAP trapper;  
Mus musculus (house mouse)

## ORGANISM

Mus musculus

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

## JOURNAL

99279253

## MEDLINE

## PUBMED

10349636

## REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

## TITLE

20499374

## JOURNAL

11042159

## MEDLINE

## PUBMED

11042159

## REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Taehiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hazada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

## TITLE

20530913

## JOURNAL

11076861

## MEDLINE

## PUBMED

11076861

## REFERENCE

4

The RIKEN Genome Exploration Research Group Phase II Team and the  
PANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

## TITLE

5

The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

## JOURNAL

6 (bases 1 to 3161)

## REFERENCE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ono, M., Ohata, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

## TITLE

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.jp,  
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

## JOURNAL

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/  
Location/Qualifiers

## FEATURES

1. 3161  
/organism="Mus musculus"

## SOURCE

Mus musculus (house mouse)

```

/mol_type="mRNA"
/db_xref="PANTOM_DB:6030436P19"
/db_xref="taxon:10090"
/clone="6030436P19"
/sex="male"
/class_type="leucis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="13 days embryo"
1..3161
/mloc_feature
/note="ANTHAX TOXIN RECEPTOR PRECURSOR (TUMOR ENDOTHELIAL
MARKER 8) homolog (mus musculus) (SWISSPROT Q99C52,
evidence: PASTY, 100%ID, 84.6%length, match=1428)"
ORIGIN
Query Match      77.3%; Score 853.4; DB 3; Length 3161;
Best Local Similarity 87.3%; Pred. No. 9.4e-242;
Matches 935; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 24 CTTGGGCGATCGGCTTCCAGTGGCTCTTTGGCCACTGTGCTCATCTGCGCCGGCA 83
Db 341 CCTGGGTGCGGGGCTGCGGGGACTCTGGGTGGCTGCACTGCTGCTGCTGCGCCGCA 400
Qy 84 AGGGGGACGCGAGGATGGGGGTCCAGCTGCTACGCGCGGATTTGACTGTAATTAT 143
Db 401 CGGGGGCGCGCGGAGATGGGGGACCACTTGTCTACGAGGATTCGACTCTACTTAT 460
Qy 144 TTGGACAAATCAGGAAGTGTGTGACACCACTGGAATCTATTCTTTGTGGAACA 203
Db 461 CTTGGACAAATCAGGAAGTGTGTGACACCACTGGAATCTATTCTTTGTGGAACA 520
Qy 204 GTTGGCTCAAAATTCATCAGCCCACTGGAATCTTTATTTGTTTCTCCACCG 263
Db 521 GTTGGCTCAATGATTCATCAGCCCACTGGAATCTTTATTTGTTTCTTTACTCG 580
Qy 264 AGGAACAACCTTATGAAACTGACAGAAAGACAAATCCCTCAAGCCCTTGAAGA 323
Db 581 AGGAACAACCTTATGAAACTGAGAGACAGGAACAGATCCGACAAAGCCCTGAAGA 640
Qy 324 ACTCCAGAAAGTTTGCAGAGAGAGACACTTAATCATGATGAAGATTTGAAGGGCCAG 383
Db 641 GCTCCAGAAAGTTTGCAGAGAGAGACACTTAATCATGATGAAGATTTGAAGGGCCAG 700
Qy 384 TGACAGATTATATGAAAAAGACAGAGGTACAGACAGCCGCTCATCTGCTTT 443
Db 701 TGACAGATTATATGAAAAAGACAGAGGTACAGACAGCCGCTCATCTGCTTT 760
Qy 444 GACTGATGAGAACTCCATGAAGATCTCTTTTCTATTGAGAGAGGCTAATAGTGC 503
Db 761 GACGAGATGGGAGCTGACAGAGACCTCTCTTACTCAAGAGAGGAGCTTAACGATC 820
Qy 504 TCGAGATCTGGTGCATATGTTTATCTGTGTGTGTAAGAAATTTCAATGACACAGCT 563
Db 821 CCGAGACCTTGGTGCATATGTTTATCTGTGTGTGTAAGAAATTTCAATGACACAGCT 880
Qy 564 GCGCCGAGATTGCGGACAGTAAGATCAATGTGTTTCCGTAATGACGGCTTTCAGGCTCT 623
Db 881 GGCTCGGATTGCAAGTAAGACAGTGTCTTCTGTGAAGACAGGCTTTCAGGCTCT 940
Qy 624 GCAAGGCACTCATCTCAATTTTGAAGAAGTCTGCACTGAATTTCTAGACGTTAAC 683
Db 941 CCAAGGCACTTATCACTCAATTTTGAAGAAGTCTGCACTGAATTTCTAGACGTTAAC 1000
Qy 684 ATCCACCATATGTGACAGAGAGTCAATTCAGTGTGTGAGAGAAACGGCTTCCGACA 743
Db 1001 ATCCACCATATGTGACAGAGAGTCAATTCAGTGTGTGAGAGAAATGCTTCCGACA 1060
Qy 744 TGCCCGCAACGTGACAGAGGTCTCTGCAAGTTCAGATCAATGCTGCTCACTCAA 803
Db 1061 TGCCCGCAACGTGACAGAGGTCTCTGCAAGTTCAGATCAATGCTGCTCACTCAA 1120
Qy 804 TGAAGAGCCCTTTTCTGTGGAAGACATTATTTACTGTGTCCAGGCGCTTCTTAAAGA 863

```

```

Db 1121 TGAGAGCCCTTCTGTGGAAGACATTATTTCTGTGCTGCCAGACCAATCTGAAAGA 1180
Qy 864 AGTTGGATGAAGACCTCACTCCAGTACAGATGAAGATGCGCTCTTTATCTCCAG 923
Db 1181 AGTTGGATGAAGACCTCACTCCAGTACAGATGAAGATGCGCTCTTTATCTCCAG 1240
Qy 924 TTCTGTATCATCAACACACACACTGTTGTGACGTTTCCATCTGGCCATCGCCTGCT 983
Db 1241 TTCTGTATCATCAACACACACACTGTTGTGACGTTTCCATCTGGCCATCGCCTGCT 1300
Qy 984 GATCTGTCTGTCTCTAGCCGCTGCTCTGTGAGTGTTCGGCCCTCTGCTGAC 1043
Db 1301 GGTCTCTCTGTCTGTGCGCCGCTGCGCTGCTGTGAGTGTTCGGCCCTCTGCTGAC 1360
Qy 1044 TGTGATTATCAAGAGGTCTCCACCCCTGCGGAGAGAGTGAAGAAA 1094
Db 1361 AGTATATCAAGAGGTCTCCACCCCTGCGGAGAGAGTGAAGAAA 1411

RESULT 3
AL542724 1006 bp mRNA linear EST 24-MAR-2004
LOCUS AL542724
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION AL542724
VERSION AL542724.3 GI:45718299
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1006)
AUTHORS L.M.B., Gruber,C., Jessee,J., and Polayres,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:30548161.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 1660.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?b=CS0DB011CF050P1&c=1660.f.
Location/Qualifiers
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB011YL09"
/class_type="PLACENTA"
/issue_type="PLACENTA"
/clone_lib="Homo sapiens PLACENTA"
/note="Vector: pCMVSPORT 6, 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched
double-strand cDNA was digested with Not I and EcoRV sites of the
pCMVSPORT 6 vector. Library was not normalized."
ORIGIN
Query Match      62.4%; Score 689.2; DB 1; Length 1006;
Best Local Similarity 99.0%; Pred. No. 4.3e-193;
Matches 703; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 1 ATGGCCACGCGGAGCGAGAGGCTTGGCATGCGCTTCCAGTGTCTCTTTGGCCACT 60
Db 293 ATGGCCACGCGGAGCGAGAGGCTTGGCATGCGCTTCCAGTGTCTCTTTGGCCACT 352
Qy 61 CTGGTGTCTATCTGCGCGGAGAGGCGAGCGAGAGATGGGGTCCAGCTGTAC 120

```

Db 353 CTGGTCTCATCTGCGCGGCAAGGGGACGACAGAGAGATGGGGGTCCAGCTCTGTAC 412  
Qy 121 GCGGATTTTGAACCTGTACTTTTGGACAAATCAGAAAGTGCTGCACCACTGGAAT 180  
Db 413 GCGGATTTTGAACCTGTACTTTTGGACAAATCAGAAAGTGCTGCACCACTGGAAT 472  
Qy 181 GAAATCTATTACTTTGGGAAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 240  
Db 473 GAAATCTATTACTTTGGGAAACAGTTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 532  
Qy 241 TCTTTTATTTGTTTCTCCACCGAGAAACAATTTAATGAAATGACAGAAACAGAA 300  
Db 533 TCTTTTATTTGTTTCTCCACCGAGAAACAATTTAATGAAATGACAGAAACAGAA 592  
Qy 301 CAAATCGTCAAGGCTTGAAGAACTCCAGAAATGTCGCGAGGAGACACTTACATG 360  
Db 593 CAAATCGTCAAGGCTTGAAGAACTCCAGAAATGTCGCGAGGAGACACTTACATG 652  
Qy 361 CATGAGGATTTGAAAGGCGCAGTGAAGCAATTTTATGAAAACAGCAAGGTTACAG 420  
Db 653 CATGAGGATTTGAAAGGCGCAGTGAAGCAATTTTATGAAAACAGCAAGGTTACAG 712  
Qy 421 ACAGCCAGCGTCAATCTTTTGAATGAGAAATCCATGAAATCTCTTTTCTAT 480  
Db 713 ACAGCCAGCGTCAATCTTTTGAATGAGAAATCCATGAAATCTCTTTTCTAT 772  
Qy 481 TCAGAGGAGGAGGCTTAATAGGCTCGAATCTGTGTGCAATGTTTACTGTGTGTG 540  
Db 773 TCAGAGGAGGAGGCTTAATAGGCTCGAATCTGTGTGCAATGTTTACTGTGTGTG 832  
Qy 541 AAAGATTTCAATGAGACACAGCTGCGCGAGTTGCGGACAGTAAGATCATGTCTTCCC 600  
Db 833 AAAGATTTCAATGAGACACAGCTGCGCGAGTTGCGGACAGTAAGATCATGTCTTCCC 892  
Qy 601 GTGAATGAGCGCTTTCAGGCTCTGCAAGCATCATCCACTCAATTTTGAAGAAGTCTGCC 660  
Db 893 GTGAATGAGCGCTTTCAGGCTCTGCAAGCATCATCCACTCAATTTTGAAGAAGTCTGCC 952  
Qy 661 ATCGAATTTCTAGAGAGTGAACCATCCACATATGTGAGAGAGTCAAT 710  
Db 953 ATCGAATTTCTAGAG--CTGAACATCCACATATGTGAGAGAGTCAAT 1000

RESULT 4  
BI823853 878 bp mRNA linear EST 04-OCT-2001  
LOCUS 60303803.F1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5179974 5',  
DEFINITION mRNA sequence.  
ACCESSION BI823853  
VERSION BI823853.1 GI:15935416  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 878)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LHAM1448 row: n column: 07  
High quality sequence stop: 804.  
Location/Qualifiers  
1..878  
FEATURES  
Source  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5179974"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_MGC\_115"  
/note="Organ: pooled brain, lung, testis; Vector:  
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb.  
Insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 62.0%; Score 684.6; DB 4; Length 878;  
Best local similarity 95.1%; Pred. No. 9,4e-192;  
Matches 783; Conservative 0; Mismatches 29; Indels 11; Gaps 7;  
Qy 1 AAGGCCACGGGAGGAGGAGGCGCTCGGATGGCTTCCAGTGGCTCTTTGGCCACT 60  
Db 57 AAGGCCACGGGAGGAGGAGGCGCTCGGATGGCTTCCAGTGGCTCTTTGGCCACT 116  
Qy 61 CTGATGCTCATCTGCGCCGGGCAAGGGGACGAGGAGAGATGGGGTCCAGCTGTCTAC 120  
Db 117 CTGATGCTCATCTGCGCCGGGCAAGGGGACGAGGAGAGATGGGGTCCAGCTGTCTAC 176  
Qy 121 GCGGATTTGACCTGTACTTCAATTTTGAACAATCAGAAAGTGTCTGCACACTGGAAT 180  
Db 177 GCGGATTTGACCTGTACTTCAATTTTGAACAATCAGAAAGTGTGTGCACTGGAAT 236  
Qy 181 GAAATCTATTACTTTGGGAAACAGTTGGCTCAAAATTCATCGAGCCACAGTTGAAGATG 240  
Db 237 GAAATCTATTACTTTGGGAAACAGTTGGCTCAAAATTCATCGAGCCACAGTTGAAGATG 296  
Qy 241 TCTTTTATTTGTTTCTCCACCGAGAAACAATTTAATGAAATGACAGAAACAGAA 300  
Db 297 TCTTTTATTTGTTTCTCCACCGAGAAACAATTTAATGAAATGACAGAAACAGAA 356  
Qy 301 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCTGCGAGAGAGACACTTACATG 360  
Db 357 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTTCTGCGAGAGAGACACTTACATG 416  
Qy 361 CATGAGGATTTGAAAGGCGCAGTGAAGCAATTTATGAAAACAGCAAGGTTACAGG 420  
Db 417 CATGAGGATTTGAAAGGCGCAGTGAAGCAATTTATGAAAACAGCAAGGTTACAGG 476  
Qy 421 ACAGCCAGCGTCAATCTTTTGAATGAGAAATCCATGAAATCTCTTTTCTAT 480  
Db 477 ACAGCCAGCGTCAATCTTTTGAATGAGAAATCCATGAAATCTCTTTTCTAT 536  
Qy 481 TCAGAGGAGGAGGCTTAATAGGCTCGAATCTGTGTGCAATGTTTACTGTGTGTG 540  
Db 537 TCAGAGGAGGAGGCTTAATAGGCTCGAATCTGTGTGCAATGTTTACTGTGTGTG 595  
Qy 541 AAAGATTTCAATGAGACACAGCTGCGCGAGTTGCGGACAGTAAGATCATGTGTCTCC 600  
Db 596 AAAGATTTCAATGAGACACAGCTGCGCGAGTTGCGGACAGTAAGATCATGTGTCTCC 655  
Qy 601 GTGAATGAGCGCTTTCAGGCTCTGCAAGCATCATCCAAATTTTG--AAGAAGTCTG 659  
Db 656 GTGAATGAGCGCTTTCAGGCTCTGCAAGCATCATCCAAATTTTGAAAGAAGTCTG 715  
Qy 660 CATC-GAAATTTCAAGCTGGAACCATCCACATATGTGAGAGAGATC-ATTCAAGTT 717  
Db 716 CATCGAAATTTCAAGCTGGAACCATCCACATATGTGAGAGAGATCATTCAAGTT 775  
Qy 718 GTGTGTGAGG--AAAGCGCTCGACATGCGCG--CAAGTGAACAGG--TCTCTGTG 770  
Db 776 GTGTGTGAGGAAACGGGTTTCCGACATGCGCGCAAAAGTGAACAGGCTCTTGTGG 835

Db	7711CAGCTTCAAGATCAATGACCTGGCTACACTCAATGAGAACGCC	813
936	AGGTTCCAGATCAATGACCTGGCTACACTCAATGAGAACGCC	878
Db	7711CAGCTTCAAGATCAATGACCTGGCTACACTCAATGAGAACGCC	813
RESULT 5	AK013005	1614 bp mRNA linear HTC 03-APR-2004
LOCUS	AK013005	Mus musculus 10, 11 days embryo whole body cDNA, RIKEN full-length
DEFINITION	enriched library, clone:281045N18 product:ANTPAX TOXIN RECEPTOR	
	PRECURESOR (TUMOR ENDOTHELIAL MARKER 8) homolog [Mus musculus], full	
	length sequence.	
ACCESSION	AK013005	
VERSION	AK013005.1 GI:12850099	
KEYWORDS	HTC; CAP trapper.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
AUTHORS	High-efficiency full-length cDNA cloning	
TITLE	Meth. Enzymol. 303, 19-44 (1999)	
JOURNAL	99279253	
MEDLINE	10349636	
PUBMED	10349636	
REFERENCE	2	
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
TITLE	High-efficiency full-length cDNA cloning	
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	99279253	
PUBMED	10349636	
REFERENCE	3	
AUTHORS	Komoto, H., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,	
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,	
JOURNAL	Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,	
MEDLINE	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, H.,	
PUBMED	Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,	
REFERENCE	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,	
AUTHORS	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	
TITLE	RIKEN Integrated sequence analysis (RISA) system-384-format	
JOURNAL	sequencing pipeline with 384 multicapillary sequencer	
MEDLINE	Genome Res. 10 (11), 1757-1771 (2000)	
PUBMED	11076661	
REFERENCE	4	
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the	
TITLE	PANTOM Consortium.	
JOURNAL	Functional annotation of a full-length mouse cDNA collection	
MEDLINE	Nature 409, 685-690 (2001)	
PUBMED	5	
REFERENCE	The PANTOM Consortium and the RIKEN Genome Exploration Research	
AUTHORS	Group Phase I & II Team.	
TITLE	Analysis of the mouse transcriptome based on functional annotation	
JOURNAL	of 60,770 full-length cDNAs	
MEDLINE	Nature 420, 563-573 (2002)	
PUBMED	6 (bases 1 to 1614)	
REFERENCE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Kono, H.,	
AUTHORS	Arizawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,	
TITLE	Furuno, M., Hanehaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,	
JOURNAL	Hiraoaka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,	
MEDLINE	Kobayashi, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Konda, M.,	
PUBMED	Koya, S., Kurikawa, C., Matsuyama, T., Miyazaki, A., Nishi, K.,	
REFERENCE	Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Oka, C.,	
AUTHORS	Saito, H., Saito, R., Sakai, K., Sakai, K., Sano, H., Sasaki, D.,	
TITLE	Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,	
JOURNAL	Shukui, H., Tsgami, M., Tagawa, A., Takahashi, F., Tanaka, T.,	
MEDLINE	Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,	
PUBMED	Yoshino, M., Muramatsu, M., and Hayashizaki, Y.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (10-JUL-2000) Yoshinori Hayashizaki, The Institute of	
TITLE	Physical and Chemical Research (RIKEN), Laboratory for Genome	
JOURNAL	Physical and Chemical Research (RIKEN), Laboratory for Genome	

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suenihiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: suenitc@res.gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216]

Please visit our web site (<http://genome.gsc.riken.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAAGATCCAGACGGCTCTTTTCTTTTCTTTTAA 3'], cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adpctccac of sequence [5' GAGAGAGAAGATTCGAGTTAATTAAATTATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SacI. Cloning sites, 5' end: XhoI; 3' end: SacI. Host: SOLR.

location/Qualifiers  
1..1614  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM.DB:2810405N18"  
/db\_xref="taxon:10090"  
/clone="2810405N18"  
/tissue\_type="whole body"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10, 11 days embryo"  
11..1306  
/note="unnamed protein product; ANTRAX TOXIN RECEPTOR PRECURSOR (TUMOR ENDOTHELIAL MARKER 8) homolog [mus musculus] (SWISSPROT|Q9CZ52, evidence: FASTA, 100%ID, 84.%length, match=1428)  
putative"  
/codon\_start=1  
/protein\_id="BAB28591.1"  
/db\_xref="GI:12850100"  
/translation="MKLTEDREBOIQROGLEBIQKTLPGSDTYMHEGFEPASQIIYENS QGYTVASVIALTDGELEHDFPYSEREARNSRLGAIVYCVGVDFNETQLAIASD KDHFVPVDFQALOGIILHSILSKSEIEILAABESTICAGESPVVVGRGNFRAIRNV DRVCSFRINDSVTLNKRPVEDVTYLCPPIKEVMGAPLOVNMNDGPSIFSSSS IITTHCSDSISLAIALVALFLALLALIMWPMPLCTVIKTPVPRESSDEDD GLPKKKPTPDASTYIGRGVGCGITRMETVRNGEKSTEGAKLEAKARVKMPQEOTYE FPESRNKNNNRRSPRSPPKMYSP1KGDLDAVLWLIRGYDRVSVMRPQDPDTRRGWR LTIICSGKHVHPGHHDGPDKPEPLTKOAMFSSFLERAFO"

polyA\_signal  
1592..1597  
/note="putative"  
1614  
/note="putative"  
/note="putative"

Query Match 61.7%; Score 681.4; DB 3; Length 1614;  
Best Local Similarity 89.0%; Pred. No. 1,1e-130;  
Matches 736; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

ORIGIN  
268 ACAACCTTAATGAAAACGTACAGAAAGACAAGAAACAAATCCGCTCAAGGCCTTAGAAGAAC TC 327  
Db 2 ACAACTTAAATGAACATACTGAGAGACAGGGAACAGATCCGACAAAGGCCTTAGAAGAGCT C 61  
  
328 CAGAAAGTTCGCAGAGAGAGAGACCTTACATGCGATGAAGATTTTGAAAGGCCAGAGAG 387  
Db 62 CAGAAAGTTCGCAGAGAGAGAGACCTTACATGCGATGAAGATTTTGAGAGGCCAGAGAG 121  
  
388 CAGATTTATTTGAAAAACAGACAGAGGGTACAGACACCAGCGTCATCATTTGCTTTGACT 447  
Db 122 CAGATTTATTTGAAACAGTCACAGAGATACAGACCGCGACGTCATCATTCGCTTTAGC 181  
  
448 GATGAGAACTCCATGAAAGATCTCTTTTCTTATTCAGAGAGAGGAGCTAATAGTCTTCCA 507

Db 182 GATGGGAGCTGCACGAGACCTCTTCTTCTACCTCAGAGGAGGCTAACCGATCCCGA 241  
Qy 508 GATCTGGTGCATTTGTTTACTGTGTGTGTGAAGATTTCAATGAGACACAGCTGGCC 567  
Db 242 GACCTTGGTGGATTGTTTACTGTGTGTGTGAAGATTTCAATGAGAACTCAATGGCT 301  
Qy 568 CGAATTCGAGACAGTGAAGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 627  
Db 302 CGAATTCGAGACAGTGAAGACACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 361  
Qy 628 GGCATCATCCCACTCAATTTTGAAGAGTCTGCATTCGAAATTTAGACGCTGAACCATCC 687  
Db 362 GGCATTCATCACTCAATTTTGAAGAGTCTGCATTCGAAATTTAGACGCTGAACCATCC 421  
Qy 688 ACCATATGTGAGAGAGATCAATTTCAAGTGTGTGTGAAGAGAAAGGCTTCGACATGCC 747  
Db 422 ACCATTCGCGGGAGAGATCTTTCAAGTGTGTGTGAAGAGAAATGGCTTCGACATGCC 481  
Qy 748 CGCAACGTGACAGGGTCTCTGACGCTTCAAGATCAATGACTGGTCACTCAATGAG 807  
Db 482 CGCAATGTGACAGGGTCTCTGACGCTTCAAAATCAATGACTGACGCTCAATGAG 541  
Qy 808 AAGCCCTTTCTGTGGAAGACATTTACTGTGTGTGAGGCGCTATCTTAAAGAAATT 867  
Db 542 AAGCCCTTTCTGTGGAAGACATTTACTGTGTGTGAGGCGCTATCTTAAAGAAATT 601  
Qy 868 GGCATGAAGGTGACCTCCAGCTCAGATGACATGACATGAGCTCTTTTATCTCCAGTTCT 927  
Db 602 GGCATGAAGGTGACCTCCAGCTCAGATGACATGACATGAGCTCTTTATCTCCAGTTCT 661  
Qy 928 GTCATCATCACCACACACACTGTTCTGACGGTTCATCTGGCCATGCTGCTGATC 987  
Db 662 GTCATCATCACCACACACACTGTTCTGACGGTTCATCTGGCCATGCTGCTGATC 721  
Qy 988 CTGTTCTGCTCTGAGCCCTGCTCTCTGAGTGTCTGAGCCCTGCTGACATGCTG 1047  
Db 722 CTCTCTGCTGCTGAGCCCTGCTCTCTGAGTGTCTGAGCCCTGCTGACATGCTG 781  
Qy 1048 ATTATCAAGAGATCTCTCCACCCCTGCGAGAGAGTGAAGAAA 1094  
Db 782 ATTATCAAGAGATCTCTCCACCCCTGCTGAGAGAGTGAAGAAA 828

RESULT 6 1054 bp mRNA linear EST 05-MAR-2002  
BM800898  
LOCUS AGENCOURT\_6420797 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5541228  
DEFINITION 5', mRNA sequence.  
ACCESSION BM800898  
VERSION BM800898.1 GI:19117721  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 1054)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCPD/DPF  
Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>  
Plate: LLM12238 row: j column: 13  
High quality sequence stop: 595.  
Location/Qualifiers  
1..1054  
/organism="Homo sapiens"

ORIGIN  
Query Match 60.0%; Score 662; DB 4; Length 1054;  
Best Local Similarity 97.5%; Pred. No. 5,4e-185;  
Matches 704; Conservative 0; Mismatches 15; Indels 3; Gaps 3;  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5541228"  
/issue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1b="NIH MGC 72"  
/note="Organ: skin; Vector: pCMV-Sport6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

Qy 1 ATGGCCACGGCGGAGCGAGAGCCCTTGCGATCGGCTTCCAGTGTCTCTTTGGCCACT 60  
Db 97 ATGGCCACGGCGGAGCGAGAGCCCTTGCGATCGGCTTCCAGTGTCTCTTTGGCCACT 156  
Qy 61 CTGGTGTCTATCTGCGCGGGCAAGGGGAGCGAGGAGGATGGGGTCCAGCTGTAC 120  
Db 157 CTGGTGTCTATCTGCGCGGGCAAGGGGAGCGAGGAGGATGGGGTCCAGCTGTAC 216  
Qy 121 GCGGATTTGACCTGTACTTCTTATTTTGAAGAAATGATGATGCTGACCACTGAAT 180  
Db 217 GCGGATTTGACCTGTACTTCTTATTTTGAAGAAATGATGATGCTGACCACTGAAT 276  
Qy 181 GAATCATTTACTTTTGTGAAGAGTGTGCTCACAAATTTCAAGCCCAAGTTGAAGATG 240  
Db 277 GAATCATTTACTTTTGTGAAGAGTGTGCTCACAAATTTCAAGCCCAAGTTGAAGATG 336  
Qy 241 TCTTTATTTTCTTCTCACCGGAGAAACCTTAATGAATGACAGAAAGAGAA 300  
Db 337 TCTTTATTTTCTTCTCACCGGAGAAACCTTAATGAATGACAGAAAGAGAA 396  
Qy 301 CAATTCGTCAAGGCCCTGAAGAACTCCAGAAATTTCTGCAGAGAGACACTTAATG 360  
Db 397 CAATTCGTCAAGGCCCTGAAGAACTCCAGAAATTTCTGCAGAGAGACACTTAATG 456  
Qy 361 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGCAAGGTAACAG 420  
Db 457 CATGAAGATTTGAAGGGCCAGTGAAGATTTATATGAAAAACAGCAAGGTAACAG 516  
Qy 421 ACAGCCAGGCTCATGATGCTTTGACTGATGAGAACTCCATGAAGATCTTTTCTAT 480  
Db 517 ACAGCCAGGCTCATGATGCTTTGACTGATGAGAACTCCATGAAGATCTTTTCTAT 576  
Qy 481 TCAGAGGAGGAGGTATAGTCTCCAGATCTTGATGCAATTTGTTACTGTGTGTG 540  
Db 577 TCAGAGGAGGAGGTATAGTCTCCAGATCTTGATGCAATTTGTTACTGTGTGTG 636  
Qy 541 AAGATTTCAATGAGACACAGCTGGCCCGGATTCGAGAGTGAAGATCAATGTTTCCC 600  
Db 637 AAGATTTCAATGAGACACAGCTGGCCCGGATTCGAGAGTGAAGATCAATGTTTCCC 696  
Qy 601 GTGAATGAGGCTTTCAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTTCC 660  
Db 697 GTGAATGAGGCTTTCAGGCTCTGCAAGGATCATCACTCAATTTTGAAGAGTCTTCC 755  
Qy 661 ATCAAAATTTCAAGAGCTGAACCATTCACCAT -ATGCA -GGAGATCAATTTCAAGTGG 718  
Db 756 ATCAAAATTTCAAGGAGGTGAACATTCCTCCATGATGTGCAAGGAGACATTTCCAGGTG 815  
Qy 719 TC 720  
Db 816 TC 817

RESULT 7 833 bp mRNA linear EST 23-JUN-2004  
LOCUS CO245219  
DEFINITION AGENCOURT\_26524177 NIH\_MGC\_212 Homo sapiens cDNA clone



ACCESSION IMAGE:30924322 5', mRNA sequence.  
VERSION C0245219  
KEYWORDS C0245219.1 GI:49108057  
SOURCE EST.  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 833)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Mary Hendrix  
CDNA Library Preparation: M. Bento Soares, University of Iowa  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: NDAm116 row: 1 column: 11  
High quality sequence start: 6  
High quality sequence stop: 688.  
Location/Qualifiers

FEATURES  
SOURCE

1..833  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30924322"  
/tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1lb="NIH\_MGC\_212"  
/note="Organ: Lung; Vector: pYX-Abc; Site\_1: EcoR I;  
Site\_2: Not I; The library was constructed according  
Bomaldi, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with Not I and then cloned  
directionally into pYX-Abc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is GATAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Query Match 59.7%; Score 658.8; DB 7; Length 833;  
Best Local Similarity 97.9%; Pred. No. 4,4e-184;  
Matches 689; Conservative 0; Mismatches 12; Indels 3; Gaps 2;  
QY 1 ATGGCCAGCGGAGCGGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTTTGGCACT 60  
DB 100 ATGGCCAGCGGAGCGGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTTTGGCACT 159  
QY 61 CTGGTGCTCATCTGCGCGGCGGAGAGCGGAGAGTGGGGGTCCAGCTGTAC 120  
DB 160 CTGGTGCTCATCTGCGCGGCGGAGAGCGGAGAGTGGGGGTCCAGCTGTAC 219  
QY 121 GCGGATTTGACCTGTACTTCTTGGCAAAATCAGAGATGTCTGACCACTGGAAT 180  
DB 220 GCGGATTTGACCTGTACTTCTTGGCAAAATCAGAGATGTCTGACCACTGGAAT 279  
QY 181 GAATCTATTACTTTGGGAAGAGTGGCTCAAAATTCATCAGCCCACTGGAATG 240  
DB 280 GAATCTATTACTTTGGGAAGAGTGGCTCAAAATTCATCAGCCCACTGGAATG 339  
QY 241 TCCCTTATTGTTTCTCAGCCGAGGAACAACCTTAATGAACGACGAGAGAGA 300  
DB 340 TCCCTTATTGTTTCTCAGCCGAGGAACAACCTTAATGAACGACGAGAGA 399

QY 301 CAATCCGTCAAGCCCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACACTTACATG 360  
DB 400 CAATCCGTCAAGCCCTTGAAGAACTCCAGAAAGTTTCCGAGAGAGACACTTACATG 459  
QY 361 CATGAAGATTTGAAGGGCAGTGAAGATTTATATGAAGAAACAGACAAAGGTTACAG 420  
DB 460 CATGAAGATTTGAAGGGCAGTGAAGATTTATATGAAGAAACAGACAAAGGTTACAG 519  
QY 421 ACAGCCAGCGTCATCTGCTTTGACGTAGAGAGAACTCCATGAAGATCTTTTCTAT 480  
DB 520 ACAGCCAGCGTCATCTGCTTTGACGTAGAGAGAACTCCATGAAGATCTTTTCTAT 579  
QY 481 TCAGAGAGGAGGCTAATAGGCTCGAGATCTTGCAATTTGTAAGTGTGTG 540  
DB 580 TCAGAGAGGAGGCTAATAGGCTCGAGATCTTGCAATTTGTAAGTGTGTG 639  
QY 541 AAATTTCAATGAGACACAGCTGGCCCGAATGGCGAGTAAAGATCATGTGTTCCC 600  
DB 640 AAATTTCAATGAGACACAGCTGGCCCGAATGGCGAGTAAAGATCATGTGTTCCC 699  
QY 601 GTGAATGACGCTTTGAGGCTCTGCAAGCATCATCTCAATTTGAAGAAGTCTGC 660  
DB 700 GTGAATGACGCTTTGAGGCTCTGCAAGCATCATCTCAATTTGAAGAAGTCTGC 759  
QY 661 ATCGAAATCTAGAGCTGAACCATCCACCATATGTGCAGAGAGA 704  
DB 760 ATCGAAATCTAGAGCTGAAC--ATCCACATATGTGCAGAGAGA 800

RESULT 8  
BE741333  
LOCUS 725 bp mRNA linear EST 15-SEP-2000  
DEFINITION 60159417P1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3948114 5',  
mRNA sequence.  
ACCESSION BE741333  
VERSION BE741333.1 GI:10155325  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 725)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: LNC610 row: n column: 19  
High quality sequence stop: 711.  
Location/Qualifiers

FEATURES  
SOURCE

1..725  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3948114"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_9"  
/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match: 57.8%; Score 637.8; DB 2; Length 725;  
 Best Local Similarity 98.8%; Pred. No. 7,3e-178;  
 Matches 653; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

```

QY 292 GACAGAGAACAAATCCGTCAGAGCCCTAGAGAACTCCAGAAAGTTCTGCCAGAGAGAC 351
DB 1 GACAGAGAACAAATCCGTCAGAGCCCTAGAGAACTCCAGAAAGTTCTGCCAGAGAGAC 60
QY 352 ACTTACATGCAATGAAAGATTGAAAAGGCCAGTACAGAGAAATTTATTAAGAAACAGACAA 411
DB 61 ACTTACATGCAATGAAAGATTGAAAAGGCCAGTACAGAGAAATTTATTAAGAAACAGACAA 120
QY 412 GGGTACAGACAGCCAGCGTCATCTGCTTTGACTGAGAGAACTCCAGAGAGATCTC 471
DB 121 GGGTACAGACAGCCAGCGTCATCTGCTTTGACTGAGAGAACTCCAGAGAGATCTC 180
QY 472 TTTTCTATTCAAGAGAGAGGCTTAATAGTCTCGAGATCTGTCATTTGTTACTGT 531
DB 181 TTTTCTATTCAAGAGAGAGGCTTAATAGTCTCGAGATCTGTCATTTGTTACTGT 240
QY 532 GTTGTGTGAAAGATTTCATAGACACAGCTGCCGGGATTTGGGACAGTAAGATCAT 591
DB 241 GTTGTGTGAAAGATTTCATAGACACAGCTGCCGGGATTTGGGACAGTAAGATCAT 300
QY 592 GTTGTTCCTGTAATGACGAGCTTTCAGGCTTCGCAAGGCATCATCAATTTTGAAG 651
DB 301 GTTGTTCCTGTAATGACGAGCTTTCAGGCTTCGCAAGGCATCATCAATTTTGAAG 360
QY 652 AAGTCTGATCATGAATTTCTAGAGCTGACATCCATATGTCAGAGAGATCTTT 711
DB 361 AAGTCTGATCATGAATTTCTAGAGCTGACATCCATATGTCAGAGAGATCTTT 420
QY 712 CAATGTGTCGAGAGAGAAAGGCTTCGACATGCGCGCAAGTGGACAGGATCTCTGC 771
DB 421 CAATGTGTCGAGAGAGAAAGGCTTCGACATGCGCGCAAGTGGACAGGATCTCTGC 480
QY 772 AGCTTCAAGATCAATGATCTCGGTCACTCAATGAGAGCCCTTTCTGTGAGAGACT 831
DB 481 AGCTTCAAGATCAATGATCTCGGTCACTCAATGAGAGCCCTTTCTGTGAGAGATCT 540
QY 832 TATTATCTGTGTCCAGCGCCCTATCTTAA-AAGAAATTGGCATGAAGTGTGACATCT 890
DB 541 TATTATCTGTGTCCAGCGCCCTATCTTAA-AAGAAATTGGCATGAAGTGTGACATCT 600
QY 891 CAGCATGAAGATGGCCCTCTTTTATCTCCAGATCTGTCAATCAACCAACACACTG 950
DB 601 CAGCATGAAGATGGCCCTCTTTTATCTCCAGATCTGTCAATCAACCAACACACTG 660
QY 951 T 951
DB 661 T 661

```

RESULT 9  
 BG326444 963 bp mRNA linear EST 27-FEB-2001  
 LOCUS 602425082P1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4553020 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG326444  
 VERSION BG326444.1 GI:13132881  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 963)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@db-remail.nih.gov  
 Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LCM1276 row: c column: 21  
 High quality sequence stop: 718.  
 Location/Qualifiers

## FEATURES

```

source
1..963
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4563020"
/tissue_type="renal cell adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 14"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

## ORIGIN

Query Match: 53.0%; Score 585; DB 4; Length 963;  
 Best Local Similarity 96.3%; Pred. No. 4.2e-162;  
 Matches 621; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

```

QY 1 ATGGCCACGGCGAGCGAGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTTTGGCCACT 60
DB 114 ATGGCCACGGCGAGCGAGAGAGCCCTCGGATCGGCTTCCAGTGGCTCTTTGGCCACT 173
QY 61 CTGCTGCTATCTGCGCGCGGCAAGGGAGCGAGAGAGATGGGGTCCAGCTGTAC 120
DB 174 CTGCTGCTATCTGCGCGCGGCAAGGGAGCGAGAGAGATGGGGTCCAGCTGTAC 233
QY 121 GGGGAAATTTGACCTGTACTTCAATTTGACAAATCAGAAATGCTGACCACTGGAAAT 180
DB 234 GGGGAAATTTGACCTGTACTTCAATTTGACAAATCAGAAATGCTGACCACTGGAAAT 293
QY 181 GAAATCTATTACTTTGTGAAACAGTGGCTCAAAATTCATCAGCCCACTTGAAGATG 240
DB 294 GAAATCTATTACTTTGTGAAACAGTGGCTCAAAATTCATCAGCCCACTTGAAGATG 353
QY 241 TCCCTTATTTGTTTCCACCCGAGAAACAACCTTAATGAACTGACAGAAACAGAA 300
DB 354 TCCCTTATTTGTTTCCACCCGAGAAACAACCTTAATGAACTGACAGAAACAGAA 413
QY 301 CAATCCGTCAGAGCCTAGAGAACTCAGAAAGTTCTGCAGAGAGACACTTACATG 360
DB 414 CAATCCGTCAGAGCCTAGAGAACTCAGAAAGTTCTGCAGAGAGACACTTACATG 473
QY 474 CATGAAGATTTGAAAGGCGCAGTGAACAATTAATATATAAACAAGGATACAGG 533
DB 534 CATGAAGATTTGAAAGGCGCAGTGAACAATTAATATATAAACAAGGATACAGG 593
QY 421 ACAAGCCAGCTCATCATTTGCTTTGACATGAGGAATCCATGAAAGATCTTTTCTAT 480
DB 534 ACAAGCCAGCTCATCATTTGCTTTGACATGAGGAATCCATGAAAGATCTTTTCTAT 593
QY 481 TCAGAGAGAGGCTAATAGGCTTCAGAGATCTTGATGTTTACTGTGTGTG 540
DB 594 TCAGAGAGAGGCTAATAGGCTTCAGAGATCTTGATGTTTACTGTGTGTG 653
QY 541 AAAGATTTCATAGACACAGCTGGCCGAGTTGGCAGTAAGATCATGTGTTCCTCC 600
DB 654 AAAGATTTCATAGACACAGCTGGCCGAGTTGGCAGTAAGATCATGTGTTCCTCC 711
QY 601 GTGAATGACGGCTTTCAGGCTTCGCAAGGATCATTCACATCAAT 645

```

```

Db              712 GTGATGACGC--TTTAAGATCTGAGAGGCATCATTCATTCTTT 754

RESULT 10
LOCUS      BG281561
DEFINITION      BG281561          820 bp      mRNA      linear      EST 21-FEB-2001
                6024024121 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4544691 5' ,
                mRNA sequence.
ACCESSION    BG281561
VERSION      BG281561.1  GI:13030486
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 820)
              NIH-MGC http://mgc.nci.nih.gov/.
              National Institutes of Health, Mammalian Gene Collection (MGC)
              Unpublished (1999)
AUTHORS      Contact: Robert Strausberg, Ph.D.
              Email: cgrabbs-remail.nih.gov
JOURNAL      Tissue Procurement: ATCC/DCTD/DTF
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNLN at: image.lnl.gov
COMMENT      Plate: LNCM128  row: h  column: 04
              High quality sequence stop: 815.
              Location/Qualifiers
FEATURES

```

```

FEATURES
SOURCE
location/Qualifiers
1. 820
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4544691"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGC 20"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adapter: GGCACGAC(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match	52.8%	Score 582.6	DB 4	Length 820
Best Local Similarity	96.1%	Prod. No. 2e-161		
Matches 683	Conservative	0	Mismatches 19	Indels 9
				Gaps 8
Qy	1	ATGGCCACGGGGGAGCGGAGACCCCTCGGCAATCGGCTTCGATGGCTCTTTGGCACT	60	
Db	112	ATGGCCACGGGGGAGCGGAGACCCCTCGGCAATCGGCTTCGATGGCTCTTTGGCACT	171	
Qy	61	CTGGTGCATCTTCGCGCGGSCAAGGGGAGCGCAGGAGATGGGGGTCCAGCTGCTAC	120	
Db	172	CTGGTGCATCTTCGCGCGGSCAAGGGGAGCGCAGGAGATGGGGGTCCAGCTGCTAC	231	
Qy	121	GGCGGATTTGACTGTACTTCAATTTGGACAATCAGGAGTGTCTGCACCACTGGAT	180	
Db	232	GGCGGATTTGACTGTACTTCAATTTGGACAATCAGGAGTGTCTGCACCACTGGAT	291	
Qy	181	GAATCTATTACTTTTGGAAACAGTT-GGCTCACAAAATTCATCAGCCCAAGTTGAGAT	239	
Db	292	GAATCTATTACTTTTGGAAACAGTTGGGTTCAAAATTCATCAGCCCAAGTTGAAAT	351	
Qy	240	GTCTCT--TTATGTTTTCTCAACCGGAGAACAACTTAATGAAACTGACGAAGACGA	297	
Db	352	GTCTCTTATATGTGTTTTCTCAACCGGAGAACAACTTAATGAAACTGACGAAGACGA	411	
Qy	298	GAAACAATCCGCTCAAGGCTGAGGAATCTCAGAAAGTTCTGCAGAGAGAGACACTTAC	357	

Db 412 GAAcAAATCCGTCAAAGcCCTAGAAgAACTCCGAAAAgTTCTCCGAGAgAGAACTTAC 471  
Qy 358 ATGcATGAAGgATTtGAAGgCCcAGTgAGcAGATTATt -ATGAAAcAGAcAAAGGTA 416  
Db 412 ATGcATGAAGgATTtGAAGgCCcAGTgAGcAGATTcATTATgAAAAcAGAcAAAGGTA 531  
Qy 417 CAGAcAGcCCAGcCGTcATtTGcCTTTgAcTgATgAGAACTcATGAAGATcTCTTTT 476  
Db 512 CAGAcAGcCCAGcCGTcATtTGcCTTTgAcTgAGAACTcATGAAGATcTCTTTGg 591  
Qy 477 CTATtCAGAGAgGgAGGcCTAAATAgGTCCTGAGATCTTgTGcAAATtGTTACTGTGTGg 536  
Db 592 CTATtCAGAGAgGgAGGcCTAAATAgGTCCTGAGATCTTgGTGcAAATtGTTACTGTGATGg 651  
Qy 537 TGTGAA -AGATTTcATATGAcAcAcAGcGTGgCCCGgATTgCGGAcAGT -AAGATcATGTGg 594  
Db 652 TGTGAACAGATTTcCATATGAcAcAcAGcGTGgCCCGgATTgCGGAcAGTAAAGATcATGTGg 711  
Qy 555 TTTCCCGGAAATGAcAGGcCTTTcAGGcCTGTGcAAAGGcATcATcCACTcAAATTTTGAAGAAG 654  
Db 712 TTTCCCGGAAATGAcAGG -TTTcAGGcCTGTGc -AGGcATcATcCACTcAAATCTTGAAGAAG 769  
Qy 655 TCCGAGATGcAAATTTcTGAAGcGTGAACcATcCACTcAAATgTGCAGAgAG 705  
Db 770 TCTCGcCTGTGAATTTcTGAAGcGTGAACcATcCA -CAATATGcACAGAcAG 819

RESULT	11
LOCUS	CFI32432
DEFINITION	641 bp mRNA linear EST 05-AUG-2003 UI-HF-FPO-awo-m-16-0-UI.r1.NH.MGC_215 Homo sapiens cDNA clone IMAGE:30560079.5', mRNA sequence.
ACCESSION	CFI32432
VERSION	CFI32432.1 GI:33215693
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 641) Bonaldo,M.F., Lennon,G. and Soares,M.B. Normalization and subtraction: two approaches to facilitate gene discovery Genome Res. 6 (9) , 791-806 (1996)
JOURNAL	
MEDLINE	97044477
PUBMED	8889548
COMMENT	Contact: Soares, MB

```

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
315 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@iowa.edu
Tissue Procurement: Mary Hendrix
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/human1.html
Seq primer: pyx-5

FEATURES
Source
Location/Qualifiers
1..641
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30560079"
/tissue_type="Chondrosarcoma Lung Metastasis cell lines"
/tab_note="DH10B (TI phage resistant)"
/clone_1ib="NH_WGC_215"
/note="Organ: Vector: pyx-Aac; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
Bontalido, Lennon and Soares, Genome Research, 6:791-806,

```

1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I and then cloned directionally into pTX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GATTAAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Query Match 50.0%; Score 552; DB 7; Length 641;  
Best Local Similarity 99.8%; Pred. No. 2,4e-152;  
Matches 563; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

1 ATGGCCACGGCGGAGGAGAGAGCCCTGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
79 ATGGCCACGGCGGAGGAGAGAGCCCTGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 138  
61 CTGGTGTCTCATCTGCGCGGCGCAAGGGGAGCAGAGGAGATGAGGAGTGGCTTCCAGCTTGTAC 120  
139 CTGGTGTCTCATCTGCGCGGCGCAAGGGGAGCAGAGGAGATGAGGAGTGGCTTCCAGCTTGTAC 198  
121 GCGCGATTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTGCTGCAACCACTGGAAT 180  
199 GCGCGATTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTGCTGCAACCACTGGAAT 258  
181 GAAATCTATTACTTGTGTGGAAGAGTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 240  
259 GAAATCTATTACTTGTGTGGAAGAGTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 318  
241 TCCTTTATTTGTTTCTCCACCCGAGAAACAATCTTAATGAAATCTGACAGAAAGAGAA 300  
319 TCCTTTATTTGTTTCTCCACCCGAGAAACAATCTTAATGAAATCTGACAGAAAGAGAA 378  
301 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTCGACGAGAGAGACCTTAACATG 360  
379 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTCGACGAGAGAGACCTTAACATG 438  
361 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTTATGAAAAAGAGAGGTTACAGG 420  
439 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTTATGAAAAAGAGAGGTTACAGG 498  
421 ACAGCCAGCGTCATCTGCTTTGACATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
499 ACAGCCAGCGTCATCTGCTTTGACATGATGAGAACTCCATGAAGATCTCTTTTCTAT 558  
481 TCAGAGAGGGAGGCTTAATAGTCTGAGATCTTGTGCAATGTTTATCTGTTGGTGTG 540  
559 TCAGAGAGGGAGGCTTAATAGTCTGAGATCTTGTGCAATGTTTATCTGTTGGTGTG 618  
541 AAAGATTTCATGAGACACAGCTG 564  
619 AAAGA-TTCAATGAGACACAGCTG 641

RESULT 12  
CF126791 780 bp mRNA linear EST 05-AUG-2003  
LOCUS CF126791  
DEFINITION UT-HF-RT0-a-v-o-15-o-UI-r1 NIH\_MGC\_214 Homo sapiens cDNA clone  
IMAGE:30560510 5', mRNA sequence.

ACCESSION CF126791  
VERSION CF126791.1 GI:33204381  
KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 780)

AUTHORS Ronald,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

## PUBMED

8889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

Tissue Procurement: Mary Hendrix  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/humanfl.html>  
Seq primer: pTX-5.

## FEATURES

## source

1..780  
location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30560510"  
/tissue\_type="Chondrosarcoma Lung Metastasis cell lines"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_1lb="NIH\_MGC\_214"  
/note="Organ: Lung; Vector: pTX-Asc; Site\_1: EcoR I;  
Site\_2: Not I; The library was constructed according  
to Soares, M. Bento Soares, University of Iowa  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with Not I and then cloned  
directionally into pTX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is GATTAAGGCCA. Tissue was provided by Mary Hendrix."

## ORIGIN

Query Match 49.0%; Score 541.4; DB 7; Length 780;  
Best Local Similarity 99.8%; Pred. No. 3.6e-149;  
Matches 542; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGGCCACGGCGGAGGAGAGAGCCCTGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 60  
238 ATGGCCACGGCGGAGGAGAGAGCCCTGGCATCGGCTTCCAGTGGCTCTCTTTGGCCACT 297  
61 CTGGTGTCTCATCTGCGCGGCGCAAGGGGAGCAGAGGAGATGAGGAGTGGCTTCCAGCTTGTAC 120  
298 CTGGTGTCTCATCTGCGCGGCGCAAGGGGAGCAGAGGAGATGAGGAGTGGCTTCCAGCTTGTAC 357  
121 GCGCGATTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTGCTGCAACCACTGGAAT 180  
358 GCGCGATTGACCTGTACTTCTTATTTGACAAATCAGAAAGTGTGCTGCAACCACTGGAAT 417  
181 GAAATCTATTACTTGTGTGGAAGAGTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 240  
418 GAAATCTATTACTTGTGTGGAAGAGTGGCTCAAAATTCATCAGCCCAAGTTGAGATG 477  
241 TCCTTTATTTGTTTCTCCACCCGAGAAACAATCTTAATGAAATCTGACAGAAAGAGAA 300  
478 TCCTTTATTTGTTTCTCCACCCGAGAAACAATCTTAATGAAATCTGACAGAAAGAGAA 537  
301 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTCGACGAGAGAGACACTTAACATG 360  
538 CAAATCGTCAAGGCTTGAAGAACTCCAGAAAGTTCGACGAGAGAGACACTTAACATG 597  
361 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTTATGAAAAAGAGAGGTTACAGG 420  
598 CATGAAGATTGAAAGGGCCAGTGAAGATTTATTTATGAAAAAGAGAGGTTACAGG 657  
421 ACAGCCAGCGTCATCTGCTTTGACATGATGAGAACTCCATGAAGATCTCTTTTCTAT 480  
658 ACAGCCAGCGTCATCTGCTTTGACATGATGAGAACTCCATGAAGATCTCTTTTCTAT 717

QY 481 TCAGAGAGGAGGCTTAATAGTCTCGAGATCTTGTCGCAATTTGTTACTGTGTGTG 540  
DB 718 TCAGAGAGGAGGCTTAATAGTCTCGAGATCTTGTCGCAATTTGTTACTGTGTGTG 777  
QY 541 AAA 543  
DB 778 AAA 780

RESULT 13  
BU131302  
LOCUS 603111791P1 CSE0CHL21 Gallus gallus cdna clone CHEST3e23 5', mRNA  
DEFINITION BU131302 823 bp mRNA linear EST 25-NOV-2002  
sequence.  
ACCESSION BU131302  
VERSION BU131302.1 GI:25343228  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianidae; Gallus.  
1 (bases 1 to 823)  
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, M.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
22335534  
MEDLINE 12445392  
PUBMED  
Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612089930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers

FEATURES  
Source  
1..823  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn, Hisex"  
/db\_xref="taxon:9031"  
/clone="CHEST3e23"  
/dev\_stage="36"  
/lab\_host="DH10B"  
/clone\_1lb="CSE0CHL21"  
/note="Organ: trunk; Vector: pBluescript II KS(+);  
Site 1: EcoRI; Site 2: NotI; Modification of pBluescript  
II KS(+) (Stratagene) vector to accommodate cDNA produced  
with the T-strimed protocol (construction of  
uni-directionally cloned cDNA libraries from messenger RNA  
for improved 3' and DNA sequencing by Glenn Fu, et al.,  
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with  
NotI and EcoRI. Ligate in double stranded adaptor  
containing BspI and BamHI sites  
[5'ggcgcgtgcagcccgacgcgacgcaaaaaag]  
[5'aattcttttttcggatcccg99gctgcagc]"

ORIGIN  
Query Match 48.4%; Score 534; DB 5; Length 823;  
Best Local Similarity 80.4%; Pred. No. 5.9e-147;  
Matches 650; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 27 CGGATGGGCTTCAGTGGCTCTCTTGGCCACTCTGTGTCTATCTGCCCGGG--CAA 84  
DB 3 CAGATTGGGCTTCAAGTGTCTTCTTTCCTTTCGACATTTTCTGTGTGTGTATGCGCAAG 62  
QY 85 GGGGAGCAGCAGGAGATGGGGGTCCAGCGCTGACGCGGATTGACCTGATCTTCAAT 144  
DB 63 TGGGAAGAGAGAGAAACGGGGGTCCTGCTGCTACGAGGGTTGATTGTACTTCAAT 122  
QY 145 TTGACAATTCAGGAGAGTGTGCTGCACCACTGAATGAATCTATTACTTTGGAAACAG 204

DB 123 CTGACAATTCAGGAGAGTGTCTGCACCACTGAGATGAATATATATCTTTGTGAAACT 182  
QY 205 TTGGCTCAAAATTCATCAGCCCAAGTTGAGATGTCCTTATTTGTTTCCACCCGA 264  
DB 183 TTGGCCGCAAGTTCTATAGCCCTCAGCTGAGATGTCCTTCACTGTTTCTCAACCA 242  
QY 265 GGAACAACTTATGAAATCGACAGAGACAGAGAACAAATCCGCTAAGGCTTGAAGAA 324  
DB 243 GGAACAACTTATGAGATTAAGAGAGACAGAGAACAGATAGCTCAGGGCTTGAAGAG 302  
QY 325 CTCAAGAAATGTTCTGCCAGAGAGACATTAATCATGATGAAGATTTGAAAGGCCA 384  
DB 303 CTTCAGAAAGTCTTCCAGAGGAGTGAACGATCATGATGAAGATTTGAAAGGCCA 362  
QY 385 GAGCAATTTATTTGAAAAAGAGAGAGGAGTGAACGAGCAGCTCATGATGCTTTG 444  
DB 363 GACAGATTTATCTATGAAATGTTCAATGTTACGAAATCGAATGTCATATTCGACTG 422  
QY 445 ACTGATGAGAACTCATGAAAGATCTCTTTTCTATTCAAGAGAGAGCTAATAGGCT 504  
DB 423 ACAGATGAGAGCTCCAGAAAGCTATTTTCTACTGAGAGAGGCCAATCGGCTCA 482  
QY 505 CGAGATCTTGGTGCATTTGTTTACTGTGTGTGTGAAGATTTCAATGACACAGCTG 564  
DB 483 AGAGCTGGAGAGCAAGATATTTGTTGTGTGAAGACTTCAATGAACCCAGCTG 542  
QY 565 GCCCGAATGGGACATTAAGATCATGTGTTCCGCTGAATGAGCGCTTCAGGCTG 624  
DB 543 GCGAATTTGCTGACACAAAGATCAATGCTTCTGTGAATGATGTTGAAGCCCTG 602  
QY 625 CAAGCATCATCCACTCAATTTTGAAGAGTCTGCATCAAAATTTAGACGTGAACCA 684  
DB 603 CAGGAGATTTACACTATTTTGAAGAAATCGATTTGAATTTGGCTGCAGAGCT 662  
QY 685 TCACCAATATGTGAGAGAGTCAATTTCAAGTGTGTGAGAGAAACGCTTCGACAT 744  
DB 663 TCAGATATATGTGAGAGAGATCAATTTCAAGTGTGTGAGAGAAATGATTTTCGACAT 722  
QY 745 GCCCGAAGCTGACAGAGGCTCTGAGCTTCAAGATCAATGATCGGTGCACCTCAAT 804  
DB 723 GCTGTAGCTGACAGAGATCTTGCAGCTTCAGATC-ATGACACAGTCACTTCAT 781  
QY 805 GAGAGCCCTTTTCTGTGAAGACACTT 832  
DB 782 GAGAGCCCTTTTGTGCGAAGAGCTTT 809

RESULT 14  
CD664532 564 bp mRNA linear EST 01-JAN-2004  
LOCUS gdb4 IMR32 SSH Homo sapiens cdna, mRNA sequence.  
ACCESSION CD664532  
VERSION CD664532.1 GI:40546157  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 564)  
De Preter, K., Patey, F., Bex, G., Menten, B., Van Roy, F., De  
Paep, A., Speleman, F. and Vandesompele, J.  
Dissection of amplicons through subtractive cDNA cloning and array  
CGH  
Unpublished (2003)  
Contact: Speleman, F.  
Center for Medical Genetics  
Ghent University Hospital  
De Pintelaan 185, 1K5, 9000 Ghent, Belgium  
Tel: 3292402451  
Fax: 3292404970  
Email: franki.speleman@ugent.be.  
Location/Qualifiers

FEATURES  
JOURNAL  
COMMENT  
AUTHORS  
TITLE  
ORIGIN

```
source
1. .564
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="Neuroblastoma"
/cell_line="IMR-92"
/lab_host="Escherichia coli"
/clone_id="IMR32 SSH"
/note="Vector: pGEM-T Easy; PCR select - suppression
subtractive hybridization (SSH)"

ORIGIN
Query Match 48.3%; Score 533.4; DB 6; Length 564;
Best Local Similarity 96.6%; Pred. No. 7.7e-147;
Matches 545; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 416 ACAGGACAGCCAGGCTCATCTTCTTGAATGAGAACTCCATGAAAGATCTCTTT 475
DB 1 ACAGGACAGCCAGGCTCATCTTCTTGAATGAGAACTCCATGAAAGATCTCTTT 60
QY 476 TCTATTCAGAGAGGAGGCTAATAGTCTGAGATCTTGTGCAATTGTTACTGTGTTG 535
DB 61 TCTATTCAGAGAGGAGGCTAATAGTCTGAGATCTTGTGCAATTGTTACTGTGTTG 120
QY 536 GTGTGAAGATTTCAATGAGACACAGCTGCCGAGATTGCGAGAGTAAGATCATGTGT 595
DB 121 GTGTGAAGATTTCAATGAGACACAGCTGCCGAGATTGCGAGAGTAAGATCATGTGT 180
QY 596 TTCCCGTGAATGAGGCTTTTCAAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAGT 655
DB 181 TTCCCGTGAATGAGGCTTTTCAAGGCTCTGCAAGGATCATCTCAATTTTGAAGAAGT 240
QY 656 CCTGCATCGAAATTTCTAGCAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAG 715
DB 241 CCTGCATCGAAATTTCTAGCAGCTGAACCATCCATATGTGAGAGAGTCAATTTCAG 300
QY 716 TTGTGCTGAGAGAAACGGCTTCGCAATGCCGCAACGTGAGAGAGGCTCTGCAAGT 775
DB 301 TTGTGCTGAGAGAAACGGCTTCGCAATGCCGCAACGTGAGAGAGGCTCTGCAAGT 360
QY 776 TCAAGATCAATGATCTCGGTCACTCAATGAGAGCCCTTTCTGTGGAAGACATTAAT 835
DB 361 TCAAGATCAATGATCTCGGTCACTCAATGAGAGCCCTTTCTGTGGAAGACATTAAT 420
QY 836 TACTGTGTCAGCCGCTTCTTAAAGAGAGTGGCATGAAGTCACTCCAGGTTCAGCA 895
DB 421 TACTGTGTCAGCCGCTTCTTAAAGAGAGTGGCATGAAGTCACTCCAGGTTCAGCA 480
QY 896 TGAACGATGCGCTCTCTTTATCTTCAGTCTGTTCATCATCAACACACACTGTTCTG 955
DB 481 TGAACGATGCGCTCTCTTTATCTTCAGTCTGTTCATCATCAACACACACTGTTCTG 540
QY 956 ACGG-TTCCATCTCGGCATGCGC 978
DB 541 ACGGATTCATCTGCGCATGCGC 564

RESULT 15
BX364961 877 bp mRNA linear EST 08-APR-2004
LOCUS BX364961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS001035YAL3 5-PRIME, mRNA sequence.
ACCESSION BX364961
VERSION BX364961.2 GI:46288818
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 877)
AUTHORS Li, W.B., Gruber, C., Jeesee, J. and Polayes, P.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
```

```
COMMENT
On May 5, 2003 this sequence version replaced gi:30370852.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gassion Creteilx, CP 5706 - 91057 EVRY cedex - FRANCE
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
1660.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?b=CS1AI009ZB070Plac=1660.f.
Location/Qualifiers
1. .877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS001035YAL3"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 48.1%; Score 531.4; DB 5; Length 877;
Best Local Similarity 94.0%; Pred. No. 3.6e-146;
Matches 548; Conservative 19; Mismatches 14; Indels 2; Gaps 2;

QY 83 AAGGGGACGCGAGGAGATGGGGGTCCAGCTCTACGCGCGGATTT-GACCTGTACTTC 141
DB 295 AAGGGGACGCGAGGAGATGGGGGTCCAGCTCTACGCGCGGATTTAKCCTGTACTTC 354
QY 142 ATTTTGAACAATACAGAAAGTGTGTCGACCACTGGAATGAATCTATTCTTTGAGAA 201
DB 355 ATTTTGAACAATACAGAAAGTGTGTCGACCACTGGAATGAATCTATTCTTTGAGAA 414
QY 202 CAGTTGGCTCAAAATTCATCAAGCCACAGTTGAGATGTCTTTATTTGTTTCTCCACC 261
DB 415 CAGTTGGCTCAAAATTCATCAAGCCACAGTTGAGATGTCTTTATTTTCTCCACC 474
QY 262 CGAGAAACAACCTTAATGAATGACAGAAAGACAGAAACAATCCGTCAAGGCTTGA 321
DB 475 CGAGAAACAACCTTAATGAATGACAGAAAGACAGAAACAATCCGTCAAGGCTTGA 534
QY 322 GAACCTCAGAAAGTCTGCCAGAGGAGACATTAATGATGAGAGATTGAAAGGAGCC 381
DB 535 GAACCTCAGAAAGTCTGCCAGAGGAGACATTAATGATGAGAGATTGAAAGGAGCC 594
QY 382 AGTAGAGCAATTTATTAATGAAGACAGAGGTAACAGACAGCCAGCATCATTTGCT 441
DB 595 AGTAGAGCAATTTATTAATGAAGACAGAGGTAACAGACAGCCAGCATCATTTGCT 654
QY 442 TTGACTGATGAGAACTCCATSAAGATCTCTTTTCTATTCTAGAGAGGCTTAATAGG 501
DB 655 TTGACTGATGAGAACTCCATSAAGATCTCTTTTCTATTCTAGAGAGGCTTAATAGG 714
QY 502 TCTCGAATCTTGATGCAATTTGTTTACTGTGTGTTGTTTCTGTTTCTGTTTCTGTTT 560
DB 715 TCTCGAATCTTGATGCAATTTGTTTACTGTGTGTTGTTTCTGTTTCTGTTTCTGTTT 774
QY 561 GCTGGCCCGGATTCGCGAGTAAGATCATGTGTGTTTCCGTAAGTACAGGCTTTGAGC 620
DB 775 GCTGGCCCGGATTCGCGAGTAAGATCATGTGTGTTTCCGTAAGTACAGGCTTTGAGC 834
QY 621 TTGCAAGGATCATCATCAATTTTGAAGAAAGTCTGCATC 663
DB 835 TTGCAAGGATCATCATCAATTTTGAAGAAAGTCTGCATC 877

Search completed: June 14, 2005, 16:43:23
```

Job time : 4263 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2005, 19:46:02 ; Search time 28.481 Seconds  
(without alignments)  
773.200 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321  
Perfect score: 1536  
Sequence: 1 GCGGRREDGPGACGFDLY.....GLSPFSSSVITTTTCSDGS 295

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	166	10.8	1155	1 US-08-286-889-46	Sequence 46, App1
2	166	10.8	1155	1 US-08-485-618-46	Sequence 46, App1
3	166	10.8	1155	1 US-08-362-652-46	Sequence 46, App1
4	166	10.8	1155	2 US-08-605-672-46	Sequence 46, App1
5	166	10.8	1155	2 US-08-482-293A-46	Sequence 46, App1
6	166	10.8	1155	2 US-08-943-363-46	Sequence 46, App1
7	166	10.8	1155	3 US-09-193-043-46	Sequence 46, App1
8	166	10.8	1155	4 US-03-688-307A-46	Sequence 46, App1
9	166	10.8	1155	4 US-03-350-259-46	Sequence 46, App1
10	166	10.8	1155	1 US-08-485-618-53	Sequence 53, App1
11	166	10.8	1155	1 US-08-362-652-53	Sequence 53, App1
12	166	10.8	1155	2 US-08-605-672-53	Sequence 53, App1
13	166	10.8	1155	2 US-08-482-293A-53	Sequence 53, App1
14	166	10.8	1155	2 US-08-943-363-53	Sequence 53, App1
15	166	10.8	1155	3 US-09-193-043-53	Sequence 53, App1
16	166	10.8	1155	4 US-09-688-307A-53	Sequence 53, App1
17	166	10.8	1155	4 US-09-350-259-53	Sequence 53, App1
18	159.5	10.4	1151	1 US-08-286-889-37	Sequence 37, App1
19	159.5	10.4	1151	1 US-08-485-618-37	Sequence 37, App1
20	159.5	10.4	1151	1 US-08-362-652-37	Sequence 37, App1
21	159.5	10.4	1151	2 US-08-605-672-37	Sequence 37, App1
22	159.5	10.4	1151	2 US-08-482-293A-37	Sequence 37, App1
23	159.5	10.4	1151	2 US-08-943-363-37	Sequence 37, App1
24	159.5	10.4	1151	3 US-09-193-043-37	Sequence 37, App1
25	159.5	10.4	1151	4 US-09-688-307A-37	Sequence 37, App1
26	159.5	10.4	1151	4 US-09-350-259-37	Sequence 37, App1
27	159.5	10.4	1151	1 US-08-485-618-55	Sequence 55, App1

28	159.5	10.4	1161	1 US-08-362-652-55	Sequence 55, App1
29	159.5	10.4	1161	2 US-08-605-672-55	Sequence 55, App1
30	159.5	10.4	1161	2 US-08-482-293A-55	Sequence 55, App1
31	159.5	10.4	1161	2 US-08-943-363-55	Sequence 55, App1
32	159.5	10.4	1161	3 US-09-193-043-55	Sequence 55, App1
33	159.5	10.4	1161	4 US-09-688-307A-55	Sequence 55, App1
34	159.5	10.4	1161	4 US-09-350-259-55	Sequence 55, App1
35	156	10.2	1161	1 US-08-173-497-2	Sequence 2, App1
36	156	10.2	1161	1 US-08-286-889-2	Sequence 2, App1
37	156	10.2	1161	1 US-08-485-618-2	Sequence 2, App1
38	156	10.2	1161	1 US-08-485-618-99	Sequence 99, App1
39	156	10.2	1161	1 US-08-362-652-2	Sequence 2, App1
40	156	10.2	1161	2 US-08-605-672-2	Sequence 2, App1
41	156	10.2	1161	2 US-08-482-293A-2	Sequence 2, App1
42	156	10.2	1161	2 US-08-943-363-2	Sequence 2, App1
43	156	10.2	1161	2 US-08-482-293A-99	Sequence 99, App1
44	156	10.2	1161	2 US-08-943-363-2	Sequence 2, App1
45	156	10.2	1161	2 US-08-943-363-99	Sequence 99, App1

## ALIGNMENTS

RESULT 1  
US-08-286-889-46  
Sequence 46, Application US/08286889  
Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,889  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: P38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULAR TYPE: protein  
US-08-286-889-46

Query Match 10.8%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
11 PACYG-GFDLYPFLDAGSV-LHHNWEIYFVQLAHKFISPOLRMSFVFSRGTTLMK 68

Db 144 PEPGQEMDIAFLIDSGSIDSDFTQMKDFVAKLMQLASTSTSFSLMYSNLKTHFT 203  
Qy 69 LTER-----EQIQLGELQKVLPGDITYMHGFERASEQIYYENROGYRTA-SVIAL 122  
Db 204 PTEFKSLSPQSLVDIAVQLQ-----GLTYTASGIQKVVLELFSKNGARKSAKKILIV 258  
Qy 123 TDGELHEDLFFYSE--REANRSRLDGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKRPDLERHVIPEAEKA---GIRYALGVGDAFREPTALQELNTIGSAPSDHVF 315  
Qy 174 PVNDGFQALQGIHSLKSCIEILAEPSITCAGESFQVVVNGFRHARNVD 227  
Db 316 KVGK-FVALRSIORIOEK---IFAIEGTESRSSSPFHMSQEGFSSALSMD 364

## RESULT 2

US-08-485-618-46  
; Sequence 46, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Seear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-46

Query Match 10.8%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GPDLYFIIDKSGSV-LHNMNEIYFVEQLAHKFIISPOLMSIVSTGRTLMK 68  
Db 144 PEPGQEMDIAFLIDSGSIDSDFTQMKDFVAKLMQLASTSTSFSLMYSNLKTHFT 203

Qy 69 LTER-----EQIQLGELQKVLPGDITYMHGFERASEQIYYENROGYRTA-SVIAL 122  
Db 204 PTEFKSLSPQSLVDIAVQLQ-----GLTYTASGIQKVVLELFSKNGARKSAKKILIV 258  
Qy 123 TDGELHEDLFFYSE--REANRSRLDGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173  
Db 259 TDGQKRPDLERHVIPEAEKA---GIRYALGVGDAFREPTALQELNTIGSAPSDHVF 315  
Qy 174 PVNDGFQALQGIHSLKSCIEILAEPSITCAGESFQVVVNGFRHARNVD 227  
Db 316 KVGK-FVALRSIORIOEK---IFAIEGTESRSSSPFHMSQEGFSSALSMD 364

## RESULT 3

US-08-362-652-46  
; Sequence 46, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Seear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-46

Query Match 10.8%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GPDLYFIIDKSGSV-LHNMNEIYFVEQLAHKFIISPOLMSIVSTGRTLMK 68  
Db 144 PEPGQEMDIAFLIDSGSIDSDFTQMKDFVAKLMQLASTSTSFSLMYSNLKTHFT 203

Qy 69 LTER-----EQIQLGELQKVLPGDITYMHGFERASEQIYYENROGYRTA-SVIAL 122  
Db 204 PTEFKSLSPQSLVDIAVQLQ-----GLTYTASGIQKVVLELFSKNGARKSAKKILIV 258  
Qy 123 TDGELHEDLFFYSE--REANRSRLDGAIVYCVGVKD-FNE-TQLARI-----ADSKDHVF 173

Db 259 TDGKFRDPLEYRHVIPAERA---GIIRYAIQVGDAREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDGFOALOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARNVD 227  
Db 316 KVGK-FVALRSTIQIOIEK---IFAIEGTSSRSSSFQHEMSQEGSSALMSMD 364

## RESULT 4

US-08-605-672-46  
; Sequence 46, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vlieten, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-605-672-46

Query Match 10.8%; Score 166; DB 2; Length 1155;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GEFDYFLIDKSGSV-LHHMNEIYVFVEQLAHKFSIPOLRMSPIVSTGTITMK 68  
Db 144 PECCGQEMDIAFLIDGSGSIDQSPFTQMKDFVAKLMGQLASTSTSFSLMQYSNLKTHFT 203  
Qy 69 LTEDR-----EQIRQGLEBIQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVKELEPHSKNGAKRSAKKIIVI 258  
Qy 123 TDGELHEDLFYYS--REARSRDLGAIYVCVGYKD-FNE-TOLARI-----ADSKDHVF 173  
Db 259 TDGKFRDPLEYRHVIPAERA---GIIRYAIQVGDAREPTALQELNTIGSAPSQDHVF 315

Qy 174 PVNDGFOALOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARNVD 227  
Db 316 KVGK-FVALRSTIQIOIEK---IFAIEGTSSRSSSFQHEMSQEGSSALMSMD 364

## RESULT 5

US-08-482-293A-46  
; Sequence 46, Application US/08482293A  
; Patent No. 5831029  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vlieten, Monica  
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,293A  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-482-293A-46

Query Match 10.8%; Score 166; DB 2; Length 1155;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 11 PACYG-GEFDYFLIDKSGSV-LHHMNEIYVFVEQLAHKFSIPOLRMSPIVSTGTITMK 68  
Db 144 PECCGQEMDIAFLIDGSGSIDQSPFTQMKDFVAKLMGQLASTSTSFSLMQYSNLKTHFT 203  
Qy 69 LTEDR-----EQIRQGLEBIQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
Db 204 FTEKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVKELEPHSKNGAKRSAKKIIVI 258  
Qy 123 TDGELHEDLFYYS--REARSRDLGAIYVCVGYKD-FNE-TOLARI-----ADSKDHVF 173  
Db 259 TDGKFRDPLEYRHVIPAERA---GIIRYAIQVGDAREPTALQELNTIGSAPSQDHVF 315  
Qy 174 PVNDGFOALOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARNVD 227

Db 316 KVGN-FVALRSIQROIQEK-----IFAIEGTESRSSSSFGHEMSQEGFSSALSMD 364

```
RESULT 6
US-08-943-363-46
; Sequence 46, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Seear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943.363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173.497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286.889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362.652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: William Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1155 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-46
```

```
Query Match 10.8%; Score 166; DB 2; Length 1155;
Best Local Similarity 28.2%; Pred. No. 6.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Cy 11 PACVY-GFDLYFIIDKSGSV-LHHMNEIYVFEQLAHKFIISPOLRMSFIVFSTGTTLMK 68
Db 144 PEPCGQEMDIAPFLIDSGSIDQSDFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203

Cy 69 LTEDR-----EQIRQGLEBELQKVLPGCDTYMHGEFPERASEQIYENRQGYRTA-SVIAL 122
Db 204 FTEFKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVLELPHSKNGAKRSACKILLVI 258

Cy 123 TDGELHEDLFFYSB--REANRSRDLAGIYVCVKD-FNE-TQLART-----ADSKDHVF 173
Db 259 TDGQKFRDPLEYRHVYIPEAEKA---GIIRYAIGVDAREPTALQELNTTIGSAPSODHVF 315

Cy 174 PVNDGFQALQGIHSLKSCIEILAAPEPTICAGSFQVAVVNGGFRHARNVD 227
Db 316 KVGN-FVALRSIQROIQEK-----IFAIEGTESRSSSSFGHEMSQEGFSSALSMD 364
```

```
RESULT 7
US-09-193-043-46
; Sequence 46, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193.043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173.497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286.889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362.652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943.363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Mus musculus
; US-09-193-043-46
```

```
Query Match 10.8%; Score 166; DB 3; Length 1155;
Best Local Similarity 28.2%; Pred. No. 6.9e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Cy 11 PACVY-GFDLYFIIDKSGSV-LHHMNEIYVFEQLAHKFIISPOLRMSFIVFSTGTTLMK 68
Db 144 PEPCGQEMDIAPFLIDSGSIDQSDFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203

Cy 69 LTEDR-----EQIRQGLEBELQKVLPGCDTYMHGEFPERASEQIYENRQGYRTA-SVIAL 122
Db 204 FTEFKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVLELPHSKNGAKRSACKILLVI 258

Cy 123 TDGELHEDLFFYSB--REANRSRDLAGIYVCVKD-FNE-TQLART-----ADSKDHVF 173
Db 259 TDGQKFRDPLEYRHVYIPEAEKA---GIIRYAIGVDAREPTALQELNTTIGSAPSODHVF 315

Cy 174 PVNDGFQALQGIHSLKSCIEILAAPEPTICAGSFQVAVVNGGFRHARNVD 227
Db 316 KVGN-FVALRSIQROIQEK-----IFAIEGTESRSSSSFGHEMSQEGFSSALSMD 364

RESULT 8
US-09-688-307A-46
; Sequence 46, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688.307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193.043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605.672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173.497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286.889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362.652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943.363
; PRIOR FILING DATE: 1997-10-03
```

NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-688-307A-46

Query Match 10.8%; Score 166; DB 4; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACG-GFDLYFLDKSGSV-LHHMNEIYFVEQLAHKFPISPOLRMSFIVSTGTTLMK 68  
DB 144 PECGQGMIDIAFLIDSGSISDQSFQMKDFVAKLMGQLASTSTSPSLMQYSNLTKHTFT 203  
QY 69 LTER-----EQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
DB 204 FTEKSSLSPOSIVDAIVQLQ-----GLTYASGIQKVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFYYSR--REANRSRLGAIYVCVGD-FNE-TOLARI-----ADSKOHVF 173  
DB 259 TDGQKFRDPLEYRHVIPAERKA---GIIRVAIGVDAREFTALQELNTIGSAPSQDHVF 315  
QY 174 PVNDGFOALQGIHSILKSCIEILAEPTICAGESFOVVVNGNFRHARNVD 227  
DB 316 KVGN-FVALRSIQRIQIOEK---IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

## RESULT 9

US-09-350-259-46  
Sequence 46, Application US/09350259  
Patent No. 6620915  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
TITLE OF INVENTION: No. 6620915el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/350.259  
EARLIER FILING DATE: 1999-07-08  
EARLIER APPLICATION NUMBER: 09/193.043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173.497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286.889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362.652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943.363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-350-259-46

Query Match 10.8%; Score 166; DB 4; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACG-GFDLYFLDKSGSV-LHHMNEIYFVEQLAHKFPISPOLRMSFIVSTGTTLMK 68  
DB 144 PECGQGMIDIAFLIDSGSISDQSFQMKDFVAKLMGQLASTSTSPSLMQYSNLTKHTFT 203  
QY 69 LTER-----EQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
DB 204 FTEKSSLSPOSIVDAIVQLQ-----GLTYASGIQKVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFYYSR--REANRSRLGAIYVCVGD-FNE-TOLARI-----ADSKOHVF 173  
DB 259 TDGQKFRDPLEYRHVIPAERKA---GIIRVAIGVDAREFTALQELNTIGSAPSQDHVF 315

QY 174 PVNDGFOALQGIHSILKSCIEILAEPTICAGESFOVVVNGNFRHARNVD 227  
DB 316 KVGN-FVALRSIQRIQIOEK---IFAIEGTESRSSSFQHEMSQEGFSSALSMD 364

## RESULT 10

US-08-485-618-53  
Sequence 53, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seair Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173.497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286.889  
FILING DATE: 5-AUG-1994  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/362.652  
FILING DATE: 21-DEC-1994  
APPLICATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-53

Query Match 10.8%; Score 166; DB 1; Length 1161;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACG-GFDLYFLDKSGSV-LHHMNEIYFVEQLAHKFPISPOLRMSFIVSTGTTLMK 68  
DB 144 PECGQGMIDIAFLIDSGSISDQSFQMKDFVAKLMGQLASTSTSPSLMQYSNLTKHTFT 203  
QY 69 LTER-----EQIRQGLEELQKVLPGGDTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
DB 204 FTEKSSLSPOSIVDAIVQLQ-----GLTYASGIQKVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFYYSR--REANRSRLGAIYVCVGD-FNE-TOLARI-----ADSKOHVF 173  
DB 259 TDGQKFRDPLEYRHVIPAERKA---GIIRVAIGVDAREFTALQELNTIGSAPSQDHVF 315  
QY 174 PVNDGFOALQGIHSILKSCIEILAEPTICAGESFOVVVNGNFRHARNVD 227

Db 316 KVG-N-FVALRSIQRIQIEK-----IFAIEGTSSRSSSFQHEMSQEGFSSALSM 364

## RESULT 11

US-08-362-652-53  
; Sequence 53, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Seer Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-53

Query Match 10.8%; Score 166; DB 1; Length 1161;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Db 11 PACYG-GFDLYFLIDSGSV-LHHNNEIYFVQQLAHKFTSPQMRMSFVSTRGTTLMK 68  
144 PECPGQEMDIAFLIDSGSIDSDPTQKDFVLMQGLASTSTSFSLMYSNLKTHFT 203  
Qy 69 LTEDR-----EQIRGLEBELQKVLPGDVTYMHGEPFRASEQIYENROGRTA-SVIAL 122  
Db 204 FTEFKSLSPQSLVDALVQLQ-----GLTYTAGSIQKVVELFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFYSE--REANRSRLGAIYVCVVKD-FNE-TOLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLRYRHVYPAEKA---GIRYALGVDAFREPTALQELNTIGSAPSODHVF 315  
Qy 174 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFOVVVANGNFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK-----IFAIEGTSSRSSSFQHEMSQEGFSSALSM 364

RESULT 12

US-08-605-672-53  
; Sequence 53, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Seer Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-605-672-53

Query Match 10.8%; Score 166; DB 2; Length 1161;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Db 11 PACYG-GFDLYFLIDSGSV-LHHNNEIYFVQQLAHKFTSPQMRMSFVSTRGTTLMK 68  
144 PECPGQEMDIAFLIDSGSIDSDPTQKDFVLMQGLASTSTSFSLMYSNLKTHFT 203  
Qy 69 LTEDR-----EQIRGLEBELQKVLPGDVTYMHGEPFRASEQIYENROGRTA-SVIAL 122  
Db 204 FTEFKSLSPQSLVDALVQLQ-----GLTYTAGSIQKVVELFHSKNGARKSAKKILIVI 258  
Qy 123 TDGELHEDLFYSE--REANRSRLGAIYVCVVKD-FNE-TOLARI-----ADSKDHVF 173  
Db 259 TDGQKFRDPLRYRHVYPAEKA---GIRYALGVDAFREPTALQELNTIGSAPSODHVF 315  
Qy 174 PVNDGFQALQGIHSLKSCIEILAAEPSTICAGESFOVVVANGNFRHARNVD 227  
Db 316 KVG-N-FVALRSIQRIQIEK-----IFAIEGTSSRSSSFQHEMSQEGFSSALSM 364

## RESULT 13

US-08-482-293A-53  
; Sequence 53, Application US/08482293A

Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5831029 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-53

Query Match 10.8%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACYG-GFDLYFIIDKSGSV-LHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMK 68  
DB 144 PECGQGMDFALIDSGSIDQSPFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203  
QY 69 LTEDR-----EQIRGLLELQKVLPGDVTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
DB 204 FTEFKSLSPQSLVDALVQLQ-----GLTYTASGIQKVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHDLFFYSR--REARRSRDLGAIYVCVGYKD-FNE-TQLARI-----ADSKDHVF 173  
DB 259 TDGQKFRDPLEYRHVYIPEAKRA--GIIRYALIGVDAREPTALQELNTIGSAPSQDHVF 315  
QY 174 PVNQGFOALOGIHSILKSCIEILAEPSITICAGESFOVVVNGNGFRHARNVD 227  
DB 316 KVGK-FVALRSIQROIOEK-----IFALGTESSRSSSFQHEMSQEGFSALSMD 364

RESULT 14  
US-08-943-363-53  
Sequence 53, Application US/08943363  
Patent No. 5831478  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael

APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5837478 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/943,363  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-363-53

Query Match 10.8%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACYG-GFDLYFIIDKSGSV-LHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMK 68  
DB 144 PECGQGMDFALIDSGSIDQSPFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203  
QY 69 LTEDR-----EQIRGLLELQKVLPGDVTYMHGFEFASQIYYENRQGYRTA-SVIAL 122  
DB 204 FTEFKSLSPQSLVDALVQLQ-----GLTYTASGIQKVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHDLFFYSR--REARRSRDLGAIYVCVGYKD-FNE-TQLARI-----ADSKDHVF 173  
DB 259 TDGQKFRDPLEYRHVYIPEAKRA--GIIRYALIGVDAREPTALQELNTIGSAPSQDHVF 315  
QY 174 PVNQGFOALOGIHSILKSCIEILAEPSITICAGESFOVVVNGNGFRHARNVD 227  
DB 316 KVGK-FVALRSIQROIOEK-----IFALGTESSRSSSFQHEMSQEGFSALSMD 364

RESULT 15  
US-09-193-043-53  
Sequence 53, Application US/09193043  
Patent No. 6251395  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica

TITLE OF INVENTION: No. 6251395e1 Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/343,363  
EARLIER FILING DATE: 1997-10-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 53  
LENGTH: 1161  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-193-043-53

Query Match 10.8%; Score 166; DB 3; Length 1161;

Best Local Similarity 28.2%; Pred. No. 6.9e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 11 PACYG-GFDLYFLIDKSGSV-LHMNEIYFVEQLAKKFIPOLRMSFIVSTRTTLMK 68  
DB 144 PECPEQEMDIAPFLIDSGSIDOSDFTQMKDFKALMGQLASTSTSFSLMOYSNLIKTHFT 203  
QY 69 LTEDR-----EQIRQGLELQKVLPGDPTVMHEGPERASBOIYYENRQGYRTA-SVITATL 122  
DB 204 FTERKSSLSFOSLVDALVQLQ-----GLTYTASGIQKVVKELFHSKNGARKSAKKILIVI 258  
QY 123 TDGELHEDLFFYSE--REANRSRDGAIYVCVGVKD-FNE-TOLARI-----ADSKDHVF 173  
DB 259 TDGQKFRDPLEYRHYTPAABKA--GIIRYALGVGDARFEPALQELNTIGSAPSQDHVF 315  
QY 174 PVNDGFQALQGIHISILKKSCEIILAAEPSTICAGESFOVVYVRNGNFRHARNVD 227  
DB 316 KVGN-FVALRSIORQIOEK---IFAIBGTSSRSSSPQHEMSQEGFSSALSMD 364

Search completed: June 13, 2005, 20:05:07  
Job time : 29.681 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 20:03:29 ; Search time 94.5881 Seconds  
(without alignments)  
1195.537 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_27\_321  
Perfect score: 1536  
Sequence: 1 GGGGRREDGGPACGFGFDLY.....GLSFSSVITTTTCHSDGS 295

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 38334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pcp.\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pcp.\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US10E\_PUBCOMB.pcp.\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US10F\_NEW\_PUB.pcp.\*
- 19: /cgn2\_6/ptodata/1/pubppaa/US11A\_PUBCOMB.pcp.\*
- 20: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pcp.\*
- 21: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pcp.\*
- 22: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536	100.0	403	11 US-09-833-245-621	Sequence 621, App
2	1536	100.0	564	10 US-09-918-715-187	Sequence 187, App
3	1536	100.0	564	10 US-09-918-715-232	Sequence 232, App
4	1536	100.0	564	14 US-10-301-822-199	Sequence 199, App
5	1536	100.0	564	16 US-10-408-765A-1823	Sequence 1823, App
6	1536	100.0	564	16 US-10-474-794-187	Sequence 187, App
7	1536	100.0	564	16 US-10-474-794-232	Sequence 232, App
8	1532	99.7	551	14 US-10-038-307-18	Sequence 18, App1
9	1532	99.7	551	14 US-10-201-292-18	Sequence 18, App1
10	1524	99.2	403	11 US-09-833-245-620	Sequence 620, App
11	1520	99.0	333	10 US-09-796-753-12	Sequence 12, App1

12	1520	99.0	333	14 US-10-038-307-2	Sequence 2, App1
13	1520	99.0	333	14 US-10-201-292-2	Sequence 2, App1
14	1520	99.0	345	14 US-10-038-307-24	Sequence 24, App1
15	1520	99.0	345	14 US-10-201-292-24	Sequence 24, App1
16	1520	99.0	564	14 US-10-038-307-20	Sequence 20, App1
17	1520	99.0	564	14 US-10-201-292-20	Sequence 20, App1
18	1519	98.9	562	10 US-09-918-715-194	Sequence 194, App
19	1519	98.9	562	10 US-09-918-715-301	Sequence 301, App
20	1519	98.9	562	16 US-10-474-794-194	Sequence 194, App
21	1519	98.9	562	16 US-10-474-794-301	Sequence 301, App
22	1507	98.1	328	14 US-10-038-307-26	Sequence 26, App1
23	1507	98.1	328	14 US-10-201-292-26	Sequence 26, App1
24	1505.5	98.0	342	14 US-10-038-307-22	Sequence 22, App1
25	1505.5	98.0	342	14 US-10-201-292-22	Sequence 22, App1
26	1500	97.7	543	14 US-10-038-307-14	Sequence 14, App1
27	1500	97.7	543	14 US-10-038-307-16	Sequence 16, App1
28	1500	97.7	543	14 US-10-201-292-16	Sequence 16, App1
29	1500	97.7	543	14 US-10-201-292-16	Sequence 16, App1
30	1490	97.0	543	14 US-10-038-307-10	Sequence 10, App1
31	1490	97.0	543	14 US-10-201-292-10	Sequence 10, App1
32	1479	96.3	534	14 US-10-038-307-12	Sequence 12, App1
33	1479	96.3	534	14 US-10-201-292-12	Sequence 12, App1
34	1419	92.4	529	14 US-10-201-292-36	Sequence 36, App1
35	1305.5	85.0	504	14 US-10-201-292-36	Sequence 36, App1
36	1178	76.7	479	14 US-10-201-292-34	Sequence 34, App1
37	1064	69.3	460	14 US-10-201-292-28	Sequence 28, App1
38	1054	68.6	460	14 US-10-201-292-30	Sequence 30, App1
39	903	58.8	538	13 US-10-047-542-99	Sequence 99, App1
40	790.5	51.5	488	10 US-09-796-753-52	Sequence 52, App1
41	790.5	51.5	488	10 US-10-038-307-6	Sequence 6, App1
42	790.5	51.5	488	14 US-10-201-292-6	Sequence 6, App1
43	790.5	51.5	488	14 US-10-368-087-16	Sequence 16, App1
44	777.5	50.6	488	15 US-10-104-047-2639	Sequence 2639, App
45	777.5	50.6	587	9 US-09-764-870-312	Sequence 312, App

ALIGNMENTS

RESULT 1  
US-09-833-245-621 Application US/09833245  
Sequence 621, Application US/09833245  
Publication No. US20040010134A1  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Albumin Fusion Proteins  
FILE REFERENCE: PFS46PCT  
CURRENT APPLICATION NUMBER: US/09/833,245  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: 60/229,358  
PRIOR FILING DATE: 2000-04-12  
PRIOR APPLICATION NUMBER: 60/256,931  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/199,384  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 2267  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 621  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-833-245-621

Query Match 100.0%; Score 1536; DB 11; Length 403;  
Best Local Similarity 100.0%; Pred. 7.4e-149; Indels 0; Gaps 0;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGRREDGGPACGFGFDLYFIIDKSGSVLHHNNELIYFYEQLAHKFISPOLHMSFVFS 60  
DB 27 GGGGRREDGGPACGFGFDLYFIIDKSGSVLHHNNELIYFYEQLAHKFISPOLHMSFVFS 86  
QY 61 TRGTTMLKLTEDDEBQIRGIEBLQKYLPGGDTYMHGFEFASQIYYENRQYRTASVII 120

Db 87 TRGTTLMKLTEDREIQRLGELQKVLPGDVTYMHGFERASEQIYYENRGYRTASVII 146  
Qy 121 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 180  
Db 147 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 206  
Qy 181 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 240  
Db 207 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 266  
Qy 241 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 295  
Db 267 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 321

## RESULT 2

US-09-918-715-187  
Sequence 187, Application US/09918715  
Publication No. US20030017157A1  
GENERAL INFORMATION:  
APPLICANT: Brad St. Croix  
APPLICANT: Bert Vogelstein  
APPLICANT: Kenneth Kinzler  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00134  
CURRENT FILING DATE: 2001-08-01  
PRIOR FILING DATE: 2001-08-01  
PRIOR FILING DATE: 2000-08-02  
PRIOR FILING DATE: 2000-08-02  
PRIOR FILING DATE: 2000-08-11  
PRIOR FILING DATE: 2000-08-11  
PRIOR FILING DATE: 2000-04-11  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 187  
LENGTH: 564  
TYPE: PR1  
ALG: IHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT  
ORGANISM: Homo sapiens  
US-09-918-715-187

Query Match 100.0%; Score 1536; DB 10; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-148;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNNEIYFYEQLAHKFISPOLRMSFIVFS 60  
Db 27 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNNEIYFYEQLAHKFISPOLRMSFIVFS 86  
Qy 61 TRGTTLMKLTEDREIQRLGELQKVLPGDVTYMHGFERASEQIYYENRGYRTASVII 120  
Db 87 TRGTTLMKLTEDREIQRLGELQKVLPGDVTYMHGFERASEQIYYENRGYRTASVII 146  
Qy 121 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 180  
Db 147 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 206  
Qy 181 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 240  
Db 207 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 266  
Qy 241 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 295  
Db 267 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 321

## RESULT 3

US-09-918-715-232  
Sequence 232, Application US/09918715  
Publication No. US20030017157A1  
GENERAL INFORMATION:  
APPLICANT: Brad St. Croix  
APPLICANT: Bert Vogelstein

APPLICANT: Kenneth Kinzler  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00134  
CURRENT FILING DATE: US/09/918, 715  
CURRENT FILING DATE: 2001-08-01  
PRIOR FILING DATE: 2001-08-01  
PRIOR FILING DATE: 2000-08-02  
PRIOR FILING DATE: 2000-08-02  
PRIOR FILING DATE: 2000-08-11  
PRIOR FILING DATE: 2000-08-11  
PRIOR FILING DATE: 2000-04-11  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 232  
LENGTH: 564  
TYPE: PR1  
ALG: IHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT  
ORGANISM: Homo sapiens  
US-09-918-715-232

Query Match 100.0%; Score 1536; DB 10; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.2e-148;  
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNNEIYFYEQLAHKFISPOLRMSFIVFS 60  
Db 27 GGGRRREDGGPACGCGFDLYFLIDKSGSVLHHNNEIYFYEQLAHKFISPOLRMSFIVFS 86  
Qy 61 TRGTTLMKLTEDREIQRLGELQKVLPGDVTYMHGFERASEQIYYENRGYRTASVII 120  
Db 87 TRGTTLMKLTEDREIQRLGELQKVLPGDVTYMHGFERASEQIYYENRGYRTASVII 146  
Qy 121 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 180  
Db 147 ALTDGELHEDLFFYSEREARNSRDLGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQ 206  
Qy 181 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 240  
Db 207 ALGGIHSILKKSCEIILAAEPSTTCAGESFOVVVRNGGFRHARNDRVLCSPKINDSVT 266  
Qy 241 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 295  
Db 267 LNEKPSVEDVTYLLCPAPILKEVGKAKALQVSMNDGSLFISSSVIITTHCSDDGS 321

## RESULT 4

US-10-301-822-199  
Sequence 199, Application US/10301822  
Publication No. US20030148410A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kametkar, Shubhangi  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
APPLICANT: Burgart, Lawrence J.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF COLON CANCER  
FILE REFERENCE: MP01-029P2RNM  
CURRENT FILING DATE: US/10/301, 822  
CURRENT FILING DATE: 2002-11-21  
PRIOR FILING DATE: US 60/339, 971  
PRIOR FILING DATE: 2001-12-10  
PRIOR FILING DATE: 2001-12-10  
PRIOR FILING DATE: 2002-03-05  
PRIOR FILING DATE: 2002-03-05  
PRIOR FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 199  
LENGTH: 564

```

! TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-199

```

Query Match	100.0%;	Score 1536;	DB 14;	Length 564;
Best Local Similarity	100.0%;	Pred. No. 1.2e-148;		
Matches 295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	GOGGREDDGPACVCGFPDLYFLIDKSGSLVHMHNIYYFVEQLAKFTSPOLBMSPIVFS	60
Db	27	GOGGREDDGPACVCGFDLYFLIDKSGSLVHMHNIYYFVEQLAKFTSPOLBMSPIVFS	86
Qy	61	TRGTTLMKLTEDREBOJROGLEBLOKVLPGDITYMEGFERASEQIYYENROGYRTASVII	120
Db	87	TRGTTLMKLTEDREBOJROGLEBLOKVLPGDITYMEGFERASEQIYYENROGYRTASVII	146
Qy	121	ALTDGELHEBLPFYSREANRSPDGLIYYCYGVDPNETGLARLADSKHVPFVNDGFO	180
Db	147	ALTDGELHEBLPFYSREANRSPDGLIYYCYGVDPNETGLARLADSKHVPFVNDGFO	206
Qy	181	ALOGIHSILIKSCIEILAAEPSTICAGESFOVVVRNGGFRHARVNDVLCSPKINDSVT	240
Db	207	ALOGIHSILIKSCIEILAAEPSTICAGESFOVVVRNGGFRHARVNDVLCSPKINDSVT	266
Qy	241	LNKRPSEVETTYLLCAPPLIKYGMKALQVSMNDGLSFISSVITTTTHCSDS	295
Db	267	LNKRPSEVETTYLLCAPPLIKYGMKALQVSMNDGLSFISSVITTTTHCSDS	321

RESULT 5  
US-10-408-765A-1823

```

1  APPLICANT:  Ghosh, Soumitra S.
2  APPLICANT:  Fany, Boi D.
3  APPLICANT:  Zhang, Bing
4  APPLICANT:  Gibson, Bradford W.
5  APPLICANT:  Taylor, Steven W.
6  APPLICANT:  Glenn, Gary M.
7  APPLICANT:  Warnock, Dale E.
8  TITLE OF INVENTION:  TARGETS FOR THERAPEUTIC INTERVENTION
9  TITLE OF INVENTION:  IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
10 FILE REFERENCE:  660088.465
11 CURRENT APPLICATION NUMBER:  US/10/408,765A
12 CURRENT FILING DATE:  2003-04-04
13 NUMBER OF SEQ ID NOS:  3077
14 SOFTWARE:  FastSeq for Windows Version 4.0
15 SEQ ID NO 1823
16 LENGTH:  564
17 TYPE:  PRF
18 ORGANISM:  Homo sapiens
19 OS-10-408-765A-1823

```

Query Match	100.0%;	Score 1536;	DB 16;	Length 564;
Best Local Similarity	100.0%;	Pred. No. 1.2e-148;		
Matches 295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	G G G R R E D G P A C Y G G F D I Y F L I D S G S V T L H M N E I Y Y P W P Q L A H K P I S P Q L R M S F V F S	60
Db	27	G G G R R E D G P A C Y G G F D I Y F L I D S G S V T L H M N E I Y Y F V S Q L A H K I S P Q L R M S F V F S	86
Qy	61	T R G T T L M K L T E D R E Q I R G L E E I O K V L P G D T Y V M H E G F E R A S E O I Y E N R G G Y T A S V I I	120
Db	87	T R G T T L M K L T E D R E Q I R G L E E I O K V L P G D T Y V M H E G F E R A S E O I Y E N R G G I T A S V I I	146
Qy	121	A L T D G E L H E D L F P Y S E R E A N R S R D L G A I V Y C V G K D F N E T Q L A R I A D S K H V P V N D G F Q	180
Db	147	A L T D G E L H E D L F E Y S E R E A N R S R D L G A I V Y C V G D F N E T Q L A R I A D S K D H V F V N D G F Q	206
Qy	181	A L O G I I H S L I L K S C E I I L A A P S T I C A G S F Q V V Y R G N G F P H A N D R V L C S P I N D S V T	240
Db	207	A L O G I I H S L I L K S C E I I L A A P S T I C A G S F Q V V Y R G N G F P H A N D R V L C S P I N D S V T	266

Qy	Db
241	267
LNKKPFSVEDTTLCPAPLIKEVGKAAQVSNNDGLSFISSSVIIITTHCSDG	LNKKPFSVEDTTLCPAPLIKEVGKAAQVSNNDGLSFISSSVIIITTHCSDG
295	321

```

RESULT 6
US-10-474-794-187
: Sequence 187, Application US/10474794
: Publication No. US2004021793A1
: GENERAL INFORMATION:
: APPLICANT: Carson-Walter, Eleanor
: APPLICANT: St. Croix, Brad
: APPLICANT: Vogelstein, Bert
: APPLICANT: Kinzler, Kenneth
: TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
: FILE REFERENCE: 1107.001'9
: CURRENT APPLICATION NUMBER: US/10/474.794
: CURRENT FILING DATE: 2003-10-14
: PRIOR APPLICATION NUMBER: 60/282,850
: PRIOR FILING DATE: 2001-04-11
: PRIOR APPLICATION NUMBER: 60/308,829
: PRIOR FILING DATE: 2001-08-01
: NUMBER OF SEQ ID NOS: 359
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 187
: LENGTH: 564
: TYPE: FRT
: ORGANISM: Homo sapiens
US-10-474-794-187

```

Query Match	100.0%;	Score 1536;	DB 16;	Length 564;
Best Local Similarity	100.0%;	Pred. No. 1.2e-148;		
Matches 295;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	GQGRREDDGPAICYGFPDIYFLIDDSGSVTHHMHNIYFVWQLAHKFTSPQLRMSFTVES	60
Dh	27	GQGRREDDGPAICYGFPDIYFLIDDSGSVTHHMHNIYFVWQLAHKFTSPQLRMSFTVES	86
Qy	61	TRGTTLMKLTEDREQIRQGLEBELQVLRPGDITVHMGEPERASEQIYYENRQGYRTASYII	120
Dh	87	TRGTTLMKLTEDREQIRQGLEBELQVLRPGDITVHMGEPERASEQIYYENRQGYRTASYII	146
Qy	121	ALTDGELHBDLPFYSREANRNRSDIQAIVYCYGVYDFNETQLAIIADSKDHVPFVNDGFO	188
Dh	147	ALTDGELHBDLPFYSREANRNRSDIQAIVYCYGVYDFNETQLAIIADSKDHVPFVNDGFO	206
Qy	181	ALQGIHSHILIKSCIEIILAAEPSTICAGESFOVVRGNGFPHANVDRVLCSPKINDSVT	240
Dh	207	ALQGIHSHILIKSCIEIILAAEPSTICAGESFOVVRGNGFPHANVDRVLCSPKINDSVT	266
Qy	241	LNKRPSPVEDTYLLCPAPILKEVGMKALQVSNMDGLSFISSYIITTHHCSDS	295
Dh	267	LNKRPSPVEDTYLLCPAPILKEVGMKALQVSNMDGLSFISSYIITTHHCSDS	321

RESULT 7  
US-10-474-794-232

APPLICANT: Carson-Walter,Eleanor  
 APPLICANT: St. Croix, Brad  
 APPLICANT: Vogelstein, Bert  
 APPLICANT: Kinzler, Kenneth  
 TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
 FILE REFERENCE: 1107, 00179  
 CURRENT APPLICATION NUMBER: US/10/474, 794  
 CURRENT FILING DATE: 2003-10-14  
 PRIOR APPLICATION NUMBER: 60/282, 850  
 PRIOR FILING DATE: 2001-04-11  
 PRIOR APPLICATION NUMBER: 60/308, 829  
 PRIOR FILING DATE: 2001-08-01

; NUMBER OF SEQ ID NOS: 359  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 232  
 ; LENGTH: 564  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-10-474-794-232

Query Match 100.0%; Score 1536; DB 16; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 1,2e-148;  
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 60  
 DB 27 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 86  
 QY 61 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 120  
 DB 87 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 146  
 QY 121 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 180  
 DB 147 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 206  
 QY 181 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 240  
 DB 207 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 266  
 QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 295  
 DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 321

## RESULT 8

US-10-038-307-18  
 ; Sequence 18, Application US/10038307  
 ; Publication No. US20030134786A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: James B. ROTTMAN  
 ; APPLICANT: Theresa L. O'KEEFE  
 ; APPLICANT: Engin OZKANAK  
 ; APPLICANT: Judith J. HEALEY  
 ; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
 ; FILE REFERENCE: 7853-253-999  
 ; CURRENT FILING DATE: US/10/038.307  
 ; CURRENT FILING DATE: 2002-06-28  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 18  
 ; LENGTH: 551  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-10-038-307-18

Query Match 99.7%; Score 1532; DB 14; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 3e-148;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 60  
 DB 27 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 86  
 QY 61 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 120  
 DB 87 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 146  
 QY 121 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 180  
 DB 147 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 206  
 QY 181 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 240  
 DB 207 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 266

QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 294  
 DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 320

## RESULT 9

US-10-201-292-18  
 ; Sequence 18, Application US/10201292  
 ; Publication No. US20030144193A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: James B. ROTTMAN  
 ; APPLICANT: Theresa L. O'KEEFE  
 ; APPLICANT: Engin OZKANAK  
 ; APPLICANT: Judith J. HEALEY  
 ; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
 ; FILE REFERENCE: 7853-253-999  
 ; CURRENT FILING DATE: US/10/201.292  
 ; CURRENT FILING DATE: 2003-02-14  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 18  
 ; LENGTH: 551  
 ; TYPE: PRF  
 ; ORGANISM: Homo sapiens  
 US-10-201-292-18

Query Match 99.7%; Score 1532; DB 14; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 3e-148;  
 Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 60  
 DB 27 GGGRRRBDGGPACGCGFDLYFLIDKSGSVLHHNMEIYFYEQLAHKFISPOLRMSFVFS 86  
 QY 61 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 120  
 DB 87 TRGTTLMKLTEDREIQROGLELOKVLPGSDTYMHGFERASQIYYENRQGYRTASVII 146  
 QY 121 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 180  
 DB 147 ALTDGELHEDLFFYSEREARNSRDGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQ 206  
 QY 181 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 240  
 DB 207 ALGGIHSILKKSCEIILAEPTTICAGESFQVVRNGFRHARNVDRVLCSEKINDSVT 266  
 QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 294  
 DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSNNDGLSFSSVIITTHCSDG 320

## RESULT 10

US-09-833-245-620  
 ; Sequence 620, Application US/09833245  
 ; Publication No. US2004010134A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Human Genome Sciences, Inc.  
 ; TITLE OF INVENTION: Albumin Fusion Proteins  
 ; FILE REFERENCE: P546PCT  
 ; CURRENT APPLICATION NUMBER: US/09/833.245  
 ; CURRENT FILING DATE: 2001-04-12  
 ; PRIOR APPLICATION NUMBER: 60/229, 358  
 ; PRIOR FILING DATE: 2000-04-12  
 ; PRIOR APPLICATION NUMBER: 60/256, 931  
 ; PRIOR FILING DATE: 2000-12-21  
 ; PRIOR APPLICATION NUMBER: 60/199, 384  
 ; PRIOR FILING DATE: 2000-04-25  
 ; NUMBER OF SEQ ID NOS: 2267  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 620  
 ; LENGTH: 403  
 ; TYPE: PRF

```

/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (1175)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (1320)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (331)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (368)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-620

Query Match          99.2%; Score 1524; DB 11; Length 403;
Best Local Similarity 99.3%; Pred. No. 1,3e-147;
Matches 293; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGRRRDGGPACGCGFDLYFLDKSGSVLHHNNEIYFVEQLAHKFTISPOLRMSFTVFS 60
DB 27 GGGRRRDGGPACGCGFDLYFLDKSGSVLHHNNEIYFVEQLAHKFTISPOLRMSFTVFS 86
QY 61 TRGTTLMKLTEDRQIRQGLELOKVLPGDVTYHGEFERASEQIYYENRQGYRTASVII 120
DB 87 TRGTTLMKLTEDRQIRQGLELOKVLPGDVTYHGEFERASEQIYYENRQGYRTASVII 146
QY 121 ALTDGELHEDLFFYSREARSRDLGAIYCVGKDPNEFQTLARIADSKDHVPVNDGFQ 180
DB 147 ALTDGELHEDLFFYSREARSRDLGAIYCVGKDPNEFQTLARIADSKDHVPVNDGFQ 206
QY 181 ALGGIHSILKSCIEILAAPESTTCAGSFQVYVVRNGGFRHARNDVRLCSFKINDSVT 240
DB 207 ALGGIHSILKSCIEILAAPESTTCAGSFQVYVVRNGGFRHARNDVRLCSFKINDSVT 266
QY 241 LNEKPSVEDVTYLLCPAPILKEVGMKALQVSNMDGSLFSSSVIITTHCSDS 295
DB 267 LNEKPSVEDVTYLLCPAPILKEVGMKALQVSNMDGSLFSSSVIITTHCSDS 321

RESULT 11
US-09-796-753-12
/ Sequence 12, Application US/09796753
/ Publication No. US20030027998A1
/ GENERAL INFORMATION:
/ APPLICANT: McCarthy, Sean A.
/ TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
/ FILE REFERENCE: 7853-227-999
/ CURRENT APPLICATION NUMBER: US/09/796,753
/ PRIOR FILING DATE: 2001-03-01
/ PRIOR APPLICATION NUMBER: 09/183,175
/ PRIOR FILING DATE: 1998-10-30
/ PRIOR APPLICATION NUMBER: 09/223,094
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/223,546
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/224,246
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/259,388
/ PRIOR FILING DATE: 1999-02-26
/ PRIOR APPLICATION NUMBER: 60/122,458
/ PRIOR FILING DATE: 1999-03-01
/ PRIOR APPLICATION NUMBER: 09/312,359
/ PRIOR FILING DATE: 1999-05-14
/ PRIOR APPLICATION NUMBER: 09/336,536
/ PRIOR FILING DATE: 1999-06-18
/ PRIOR APPLICATION NUMBER: 09/342,687
/ PRIOR FILING DATE: 1999-06-29
/ PRIOR APPLICATION NUMBER: 09/345,464
/ PRIOR FILING DATE: 1999-06-30
/ PRIOR APPLICATION NUMBER: 09/365,164
/ PRIOR FILING DATE: 1999-07-30
```

```

/ PRIOR APPLICATION NUMBER: 09/399,723
/ PRIOR FILING DATE: 1999-09-20
/ PRIOR APPLICATION NUMBER: 09/409,634
/ PRIOR FILING DATE: 1999-09-30
/ PRIOR APPLICATION NUMBER: 09/471,179
/ PRIOR FILING DATE: 1999-12-23
/ PRIOR APPLICATION NUMBER: 09/474,071
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/474,072
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/514,010
/ PRIOR FILING DATE: 2000-02-25
/ PRIOR APPLICATION NUMBER: 09/516,745
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 09/572,002
/ PRIOR FILING DATE: 2000-05-14
/ PRIOR APPLICATION NUMBER: 09/597,993
/ PRIOR FILING DATE: 2000-06-19
/ PRIOR APPLICATION NUMBER: 09/599,596
/ PRIOR FILING DATE: 2000-06-22
/ PRIOR APPLICATION NUMBER: 09/630,334
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 09/606,565
/ PRIOR FILING DATE: 2000-06-29
/ PRIOR APPLICATION NUMBER: 09/606,317
/ PRIOR FILING DATE: 2000-06-29
/ PRIOR APPLICATION NUMBER: 09/665,666
/ PRIOR FILING DATE: 2000-09-20
/ PRIOR APPLICATION NUMBER: 09/677,751
/ PRIOR FILING DATE: 2000-09-30
/ NUMBER OF SEQ ID NOS: 162
/ SEQ ID NO 12
/ LENGTH: 333
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-796-753-12

Query Match          99.0%; Score 1520; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 2.5e-147;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRDGGPACGCGFDLYFLDKSGSVLHHNNEIYFVEQLAHKFTISPOLRMSFTVFS 60
DB 27 GGGRRRDGGPACGCGFDLYFLDKSGSVLHHNNEIYFVEQLAHKFTISPOLRMSFTVFS 86
QY 61 TRGTTLMKLTEDRQIRQGLELOKVLPGDVTYHGEFERASEQIYYENRQGYRTASVII 120
DB 87 TRGTTLMKLTEDRQIRQGLELOKVLPGDVTYHGEFERASEQIYYENRQGYRTASVII 146
QY 121 ALTDGELHEDLFFYSREARSRDLGAIYCVGKDPNEFQTLARIADSKDHVPVNDGFQ 180
DB 147 ALTDGELHEDLFFYSREARSRDLGAIYCVGKDPNEFQTLARIADSKDHVPVNDGFQ 206
QY 181 ALGGIHSILKSCIEILAAPESTTCAGSFQVYVVRNGGFRHARNDVRLCSFKINDSVT 240
DB 207 ALGGIHSILKSCIEILAAPESTTCAGSFQVYVVRNGGFRHARNDVRLCSFKINDSVT 266
QY 241 LNEKPSVEDVTYLLCPAPILKEVGMKALQVSNMDGSLFSSSVIITTHCS 292
DB 267 LNEKPSVEDVTYLLCPAPILKEVGMKALQVSNMDGSLFSSSVIITTHCS 318

RESULT 12
US-10-038-307-2
/ Sequence 2, Application US/10038307
/ Publication No. US20030134786A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALRY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
```

/ CURRENT APPLICATION NUMBER: US/10/038,307  
/ CURRENT FILING DATE: 2002-06-28  
/ NUMBER OF SEQ ID NOS: 26  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 2  
/ LENGTH: 333  
/ TYPE: PRF  
/ ORGANISM: Homo sapiens  
US-10-038-307-2

Query Match 99.0%; Score 1520; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2,5e-147;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 60  
DB 27 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 120  
DB 87 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 180  
DB 147 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 206  
QY 181 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 240  
DB 207 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 266  
QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 292  
DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 318  
RESULT 13  
US-10-201-292-2  
/ Sequence 2, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 2  
/ LENGTH: 333  
/ TYPE: PRF  
/ ORGANISM: Homo sapiens  
US-10-201-292-2

Query Match 99.0%; Score 1520; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2,5e-147;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 60  
DB 27 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 120  
DB 87 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 180  
DB 147 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 206  
QY 181 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 240

DB 207 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 266  
QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 292  
DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 318

RESULT 14  
US-10-038-307-24

/ Sequence 24, Application US/10038307  
/ Publication No. US20030134786A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/038,307  
/ CURRENT FILING DATE: 2002-06-28  
/ NUMBER OF SEQ ID NOS: 26  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 24  
/ LENGTH: 345  
/ TYPE: PRF  
/ ORGANISM: Homo sapiens  
US-10-038-307-24

Query Match 99.0%; Score 1520; DB 14; Length 345;  
Best Local Similarity 100.0%; Pred. No. 2,6e-147;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 60  
DB 27 GGGRRRDEGPGACGFDLYFILDKSGSVLHNNIYFVEQLAHKFIISQLRMSFIVFS 86  
QY 61 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 120  
DB 87 TRGTTLMKLTEDRQIRQGLELQKVLPGDPTVMHEGFERASQIYYENRQGYRTASVII 146  
QY 121 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 180  
DB 147 ALTDGELHEDLFFYSEREARNSRDGAIVYCVGVKDFNEQTLARIADSKDHVPVNDGFQ 206  
QY 181 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 240  
DB 207 ALGGIHSILKSCIEILAAEPSTICAGESFQVYVNGRFRHARNVDRVLCSEFKINDSVT 266  
QY 241 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 292  
DB 267 LNEKPSVEDTYLLCPAPILKEVGMKALQVSMNDGLSFSSVITTTTHCS 318

RESULT 15  
US-10-201-292-24

/ Sequence 24, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 24  
/ LENGTH: 345  
/ TYPE: PRF  
/ ORGANISM: Homo sapiens

US-10-201-292-24

Query Match 99.0%; Score 1520; DB 14; Length 345;  
Best Local Similarity 100.0%; Pred. No. 2.6e-147;  
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 GGGRRREDGGPACYGFDLYFIIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIIFS 60
      |||||||
Db      27 GGGRRREDGGPACYGFDLYFIIDKSGSVLHHMNEIYFVEQLAHKFIISPOLRMSFIIFS 86
      |||||||
QY      61 TRGTTLMKLTEDREQIRQGLEBELQVLPGGPTVMHEGFERASEOIYYENRQGYRTASVIT 120
      |||||||
Db      87 TRGTTLMKLTEDREQIRQGLEBELQVLPGGPTVMHEGFERASEOIYYENRQGYRTASVIT 146
      |||||||
QY      121 ALTDGELHEDLFFYSEREANRSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 180
      |||||||
Db      147 ALTDGELHEDLFFYSEREANRSRDIGAIYVCVGVDFNETQLARIADSKDHVPVNDGFQ 206
      |||||||
QY      181 ALQGIHSHILKKSCTEIIAEPSTICAGESFQVVVRGNGFRHANNVDRVLCSPKINDSVT 240
      |||||||
Db      207 ALQGIHSHILKKSCTEIIAEPSTICAGESFQVVVRGNGFRHANNVDRVLCSPKINDSVT 266
      |||||||
QY      241 LNEKPFVSVEDTYLCPAPILKEVGKAKALQVSNNDGLSFISSVIITTHCS 292
      |||||||
Db      267 LNEKPFVSVEDTYLCPAPILKEVGKAKALQVSNNDGLSFISSVIITTHCS 318
      |||||||
```

Search completed: June 13, 2005, 20:36:49  
Job time : 96.5881 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 19:47:08 ; Search time 21.1647 Seconds

(without alignment)  
1341.095 Million cell updates/sec

Title: us-09-970-076-2\_COPY\_27\_321

Perfect score: 1536

Sequence: 1 GCGRRRDGPGACYGDFLY.....GLSPSSSVITTTTCSDGS 295

Scoring table: BLOSUM62

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	9.5	1153	1 RWHUB	cell surface glyco
2	143.5	9.3	1163	1 RWHUC	cell surface glyco
3	143	9.3	3124	2 A40020	collagen alpha 1(X
4	141	9.2	1153	2 S00551	leukocyte surface
5	139	9.0	724	2 A48569	antigen Em100 - E1
6	137	8.9	1747	2 A45974	collagen alpha 1(X
7	134.5	8.8	712	2 A45638	immunodominant mic
8	131	8.5	760	1 C2MS	classical-compleme
9	130	8.5	1857	2 S31212	collagen alpha 1(X
10	130	8.5	1888	2 S78476	collagen alpha 1(X
11	128	8.3	3051	2 S42373	hypothetical prote
12	118	7.7	929	2 I51027	type XII collagen
13	113.5	7.4	496	2 A37979	cartilage matrix p
14	113.5	7.4	764	1 BBHU	complement factor
15	111.5	7.3	2944	2 A54849	collagen alpha 1(V
16	110.5	7.2	761	1 BMS	complement factor
17	109.5	7.1	1029	1 S21369	collagen alpha 2(V
18	108.5	7.1	500	2 S66522	cartilage matrix p
19	107.5	7.0	493	2 A33809	cartilage matrix p
20	105.5	6.9	567	2 T38797	hypothetical prote
21	104	6.8	747	2 I51579	complement factor
22	100.5	6.5	574	2 A46283	sporozoite surface
23	100	6.5	460	2 T33087	hypothetical prote
24	99.5	6.5	597	2 S33578	rop protein - fru1
25	99	6.4	559	2 S04531	thrombospondin-rel
26	99	6.4	559	2 B96958	dnak protein (heat
27	99	6.4	817	2 T50240	kinesin-like prote
28	99	6.4	917	2 S09646	collagen alpha 2(V
29	99	6.4	1018	1 CGH02A	collagen alpha 2(V

30	98.5	6.4	537	2 T04822	hypothetical prote
31	98	6.4	741	2 T46488	hypothetical prote
32	98	6.4	3176	2 CGH03A	collagen alpha 3(V
33	97.5	6.3	642	2 H81185	dnak protein NMB05
34	97	6.3	3137	2 A37797	collagen alpha 3(V
35	96.5	6.3	676	2 T47637	hypothetical prote
36	96	6.2	272	2 A55348	hypothetical prote
37	96	6.2	689	2 F84811	integrin alpha-1 -
38	95.5	6.2	334	2 AF1166	probable retroelem
39	95.5	6.2	932	2 JC5953	transcription regu
40	95.5	6.2	1179	2 A53213	inter-alpha-inhibi
41	94.5	6.2	1151	2 A45226	integrin alpha-B c
42	94	6.1	843	2 A40970	undulin 1 - human
43	93	6.1	340	2 E70121	hypothetical prote
44	92.5	6.0	642	2 B81917	probable chaperone
45	92	6.0	292	2 B83736	transposase (04) B

#### ALIGNMENTS

##### RESULT 1

RWHUB

cell surface glycoprotein CD11b precursor [validated] - human  
N/Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac  
eukocyte integrin alpha chain; neutrophil adherence receptor alphan chain

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 09-Jul-2004

C/Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567

R/Corbl, A.L.; Kishimoto, T.K.; Miller, U.U.; Springer, T.A.

U. Biol. Chem. 263, 12403-12411, 1988

A/Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11

B.

A/Reference number: A31108; MUID:88315033; PMID:2457584

A/Accession: A31108

A/Molecule type: mRNA

A/Residues: 1-1153 <COR>

A/Cross-references: UNIPROT:P11215; GB:U03925; NID:G187284; PIDN:AAA59544.1; PID:G307148

A/Note: part of this sequence was confirmed by protein sequencing

R/Arnaut, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.

J. Cell Biol. 106, 2153-2158, 1988

A/Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac

A/Reference number: A28915; MUID:88257215; PMID:2454931

A/Accession: A28915

A/Molecule type: mRNA

A/Residues: 1-499, 501-965, 'P', 967-1153 <ARN>

A/Cross-references: GB:M18044; GB:U03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594;

A/Note: the authors translated the codon TAC for residue 1129 as Thr

A/Note: part of this sequence, including the amino end of the mature protein, was confirm

R/Shelley, C.S.; Arnaut, M.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A/Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regu

A/Reference number: A41600; MUID:92073318; PMID:1683702

A/Accession: A41600

A/Molecule type: DNA

A/Residues: 1-9 <SHB>

A/Cross-references: GB:M76724; NID:G180018; PIDN:AAA5410.1; PID:G553215

R/Arnaut, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988

A/Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesic

A/Reference number: A94193; MUID:88190151; PMID:2833753

A/Accession: A30892

A/Molecule type: mRNA

A/Residues: 917-1042 <AR2>

A/Cross-references: GB:M18044

R/Hickstein, D.D.; Hickey, M.U.; Ozols, J.; Baker, D.M.; Beck, A.L.; Roth, G.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A/Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor

A/Reference number: A32218; MUID:89088893; PMID:2563162

A/Accession: A32218

A/Molecule type: mRNA

A/Residues: 9-1153 <HIC>

A/Cross-references: GB:U04145; NID:G189068; PIDN:AAA59903.1; PID:G386975

A>Note: part of this sequence was confirmed by protein sequencing  
R: Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
J. Immunol. 150, 480-490, 1993  
A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1r  
n during evolution.  
A:Reference number: A46526; MUID:93123748; PMID:8419480  
A:Accession: A46526  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-499, 501-1153 <FLR>  
A:Cross-references: GB:SS2227; NID:g263047; PID:AA24821.1; PID:g263049  
A>Note: the last three bases of intron 13, CAG, are included in some but not all mature  
A:Note: sequence extracted from NCBI backbone (NCBI:121963)  
R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnout, M.A.  
Biochim. Biophys. Acta 874, 368-371, 1986  
A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
A:Reference number: A90664; MUID:87076671; PMID:3539202  
A:Accession: A26091  
A:Molecule type: protein  
A:Residues: 17-31 <PIE>  
A:Experimental source: granulocytes  
R:Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.  
Blood 79, 865-870, 1992  
A:Title: Characterization of the myeloid-specific CD11b promoter.  
A:Reference number: 152567; MUID:92144986; PMID:1346576  
A:Accession: 152567  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: RNA  
A:Residues: 1-9 <RSS>  
A:Cross-references: GB:M4477; NID:g180184; PID:AA51960.1; PID:g553219  
A:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
C:Genetic:  
A:Gene: GDB:ITGAM; CR3A  
A:Cross-references: GDB:120599; OMIM:120980  
A:Map position: 16p11.2-16p11.2  
A:Note: promoter containing a GATA motif and two Sp1 consensus binding sites  
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: cell surface glycoprotein CD11b #status predicted <MNT>  
F:17-1108/Domain: extracellular #status predicted <EXT>  
F:148-318/Domain: von Willebrand factor type A repeat homology <VMA2>  
F:465-473/Region: calcium/magnesium binding #status predicted  
F:530-538/Region: calcium/magnesium binding #status predicted  
F:593-601/Region: calcium/magnesium binding #status predicted  
F:1109-1134/Domain: transmembrane #status predicted <TM>  
F:1135-1153/Domain: intracellular #status predicted <INT>  
F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
Query Match 9.5%; Score 145.5; DB 1; Length 1153;  
Best Local Similarity 26.2%; Pred. No. 0.0016;  
Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;

N:Alternate names: leukocyte adhesion receptor p150,95 alpha chain  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 09-Jul-2004  
A:Accession: A36584; A35543; 500864  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 12750-12751, 1990  
A:Reference number: A36584  
A:Accession: A36584  
A:Content: extratum  
A:Molecule type: DNA  
A:Residues: 1-1163 <COR>  
A:Cross-references: UNIPROT:P20702  
A:Note: this revision to the sequence from reference A35543 includes the carboxyl end  
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2788, 1990  
A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A:Reference number: A35543; MUID:90153906; PMID:2303426  
A:Accession: A35543  
A:Molecule type: DNA  
A:Residues: 1-834 <CO2>  
A:Note: this sequence has been revised in reference A36584  
R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
EMBO J. 6, 4023-4028, 1987  
A:Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A:Reference number: 500864; MUID:816645; PMID:3327687  
A:Accession: 500864  
A:Molecule type: mRNA  
A:Residues: 1-755, 'L', 757-1163 <CO3>  
A:Cross-references: GB:M81695; EMBL:090093; NID:g487829; PID:AA59180.1; PID:g487830  
A:Note: part of this sequence was confirmed by protein sequencing  
C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my  
C:Genetic:  
A:Gene: GDB:ITGAX; CD11C  
A:Cross-references: GDB:119758; OMIM:151510  
A:Map position: 16p11.2-16p11.2  
A:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MNT>  
F:20-1107/Domain: extracellular #status predicted <EXT>  
F:149-319/Domain: von Willebrand factor type A repeat homology <VMA4>  
F:1108-1133/Domain: transmembrane #status predicted <TM>  
F:1134-1163/Domain: intracellular #status predicted <INT>  
F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status prec  
Query Match 9.3%; Score 143.5; DB 1; Length 1163;  
Best Local Similarity 24.8%; Pred. No. 0.0024;  
Matches 55; Conservative 43; Mismatches 89; Indels 35; Gaps 10;

Query Match 9.5%; Score 145.5; DB 1; Length 1153;  
Best Local Similarity 26.2%; Pred. No. 0.0016;  
Matches 60; Conservative 45; Mismatches 81; Indels 43; Gaps 12;

Query Match 9.3%; Score 143.5; DB 1; Length 1163;  
Best Local Similarity 24.8%; Pred. No. 0.0024;  
Matches 55; Conservative 43; Mismatches 89; Indels 35; Gaps 10;

R.Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shimomura, T.; Tanaka, H.; Nishida, Y.; Obata, U. Cell Biol. 115, 209-221, 1991  
 A>Title: The complete primary structure of type XII collagen shows a chimeric molecule with a novel region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.  
 A/Reference number: A40020; MUID:92011862; PMID:1918137  
 A/Accession: A40020  
 A/Molecule type: mRNA  
 A/Residues: 1-3124 <YAM>  
 A/Cross-references: UNIPROT:P13944; GB:D00824; NID:9222810; PIDN:BAA00701.1; PID:9222811  
 A/Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and, R.Gordon, M.K.; Garecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.  
 J. Biol. Chem. 264, 19772-19778, 1989  
 A>Title: Type XII collagen. A large multidomain molecule with partial homology to type I  
 A/Reference number: A34485; MUID:90062079; PMID:2584192  
 A/Accession: A34485  
 A/Molecule type: mRNA  
 A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>  
 A/Cross-references: EMBL:U05137; NID:9211284; PIDN:AAA8635.1; PID:9211285  
 A/Accession: B34485  
 A/Molecule type: protein  
 A/Residues: 2772-2792;2846-2873 <GOR2>  
 R.Gordon, M.K.; Garecke, D.R.; Olsen, B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 6040-6044, 1987  
 A>Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA  
 A/Reference number: A28037; MUID:87317590; PMID:3476925  
 A/Accession: A28037  
 A/Molecule type: mRNA  
 A/Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>  
 A/Cross-references: EMBL:M17375; NID:9211649; PIDN:AAA8718.1; PID:9211650  
 A/Note: This sequence has been revised in reference A34485  
 R.Koch, M.; Bernasconi, C.; Chiquet, M.  
 Eur. J. Biochem. 207, 847-856, 1992  
 A>Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of  
 A/Reference number: 523814; MUID:92362621; PMID:1332460  
 A/Accession: 523814  
 A/Molecule type: protein  
 A/Residues: 'X', 1333, 'Q', 1335-1347;1314-1928;2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <R>  
 J. Dublet, B.; van der Rest, M.  
 J. Biol. Chem. 267, 17724-17727, 1992  
 A>Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peptide  
 A/Reference number: 522254; MUID:88087065; PMID:3121603  
 A/Accession: 522254  
 A/Molecule type: protein  
 A/Residues: 2831-2832, 'T', 2834, 'R', 2836-2843;3002-3014 <DUB>  
 R.Tueb, J.; Tueb, B.  
 Biochim. Biophys. Acta 1171, 97-98, 1992  
 A>Title: The two splice variants of collagen XII share a common 5' end.  
 A/Reference number: 528811; MUID:93042014; PMID:1420368  
 A/Accession: 528811  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>  
 A/Cross-references: EMBL:X67327  
 C/Genetics:  
 A/Intons: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1  
 C/Keywords: alternative splicing; cell binding; cell binding; coll; connective tissue; disulfide  
 F:1-23/Domin: signal sequence #status predicted <SIG>  
 F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>  
 F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
 F:24-114/Domin: IIA #status predicted <IIA>  
 F:24-105/Domin: fibronectin type III repeat homology <FN3A>  
 F:137-301/Domin: von Willebrand factor type A repeat homology <VMA1>  
 F:133-425/Domin: IIB #status predicted <IIB>  
 F:133-414/Domin: fibronectin type III repeat homology <FN3B>  
 F:437-601/Domin: von Willebrand factor type A repeat homology <VMA2>  
 F:629-1178/Domin: IIC #status predicted <IIC>  
 F:630-711/Domin: fibronectin type III repeat homology <FN3C>  
 F:721-802/Domin: fibronectin type III repeat homology <FN3D>  
 F:812-895/Domin: fibronectin type III repeat homology <FN3E>  
 F:905-986/Domin: fibronectin type III repeat homology <FN3F>  
 F:995-1076/Domin: fibronectin type III repeat homology <FN3G>  
 F:1086-1169/Domin: fibronectin type III repeat homology <FN3H>  
 F:1197-1361/Domin: von Willebrand factor type A repeat homology <VMA3>

F:1384-2295/Domin: IID #status predicted <IID>  
 F:1384-1465/Domin: fibronectin type III repeat homology <FN3I>  
 F:1474-1571/Domin: fibronectin type III repeat homology <FN3J>  
 F:1566-1647/Domin: fibronectin type III repeat homology <FN3K>  
 F:1555-1738/Domin: fibronectin type III repeat homology <FN3L>  
 F:1756-1838/Domin: fibronectin type III repeat homology <FN3M>  
 F:1847-1928/Domin: fibronectin type III repeat homology <FN3N>  
 F:1937-2019/Domin: fibronectin type III repeat homology <FN3O>  
 F:2028-2110/Domin: fibronectin type III repeat homology <FN3P>  
 F:2119-2199/Domin: fibronectin type III repeat homology <FN3Q>  
 F:2207-2294/Domin: fibronectin type III repeat homology <FN3R>  
 F:2325-2490/Domin: von Willebrand factor type A repeat homology <VMA4>  
 F:2438-2440/Domin: cell adhesion #status predicted  
 F:2509-2750/Domin: IXP, homologous to NC4 domain of type IX collagen #status predicted  
 F:2751-2902/Domin: collagenous COL2 #status predicted <COL2>  
 F:2899-2901/Domin: cell attachment (R-G-D) motif  
 F:2903-2945/Domin: non-collagenous NC2 #status predicted <NC2>  
 F:2946-3048/Domin: collagenous COL1 #status predicted <COL1>  
 F:3049-3124/Domin: non-collagenous NC1 #status predicted <NC1>  
 F:32-1006;1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (coval)  
 F:2780,2789,2836,2842,2860,2866,2869,3004,3007/Modified site: hydroxyproline (Pro) #status  
 Query Match 9.3%; Score 143; DB 2; Length 3124;  
 Best Local Similarity 26.1%; Pred. No. 0.0098;  
 Matches 65; Conservative 51; Mismatches 97; Indels 36; Gaps 14;  
 QY 18 DLYFLDKGSV-LHHNNEIYFVEOLAHKF-ISP-OLRMSFVSTRTTLMKLTEDRE 74  
 Db 439 DVVFPLVGSISIGIANVKKRAFLFVLSFESPRKQVSLVOYSR--DPRHESLNRY 496  
 QY 75 QIROGLEBELQVLP--GGDTYMEGFERASEQIYENRQGYR--TASVIALTDGLHED 130  
 Db 497 NRVVDIIQALINTFPRYGSINTGKAMTYVEKVFVTSK--GSRPVPRVMTLITDGK--SSD 554  
 QY 131 LFFYSREARSRDGLAVYCVGKDPNEQLRIAD--SKDRVPVNDFOQLGIHS 188  
 Db 555 AF--KSPAKLRADVEIYFAVGKQAVRTELEIASPPAEHYVTEYD--FDLQRIISFE 610  
 QY 189 ILKSCIEIIAEPSTICAGESFQVVRNGGFRHARV--DRYLCSEKINDSVTLNEKP 245  
 Db 611 LTQSVLCIRT-----EQELAIRKSYPAKRMVSDVTSDFKWSMAGSE-- 657  
 QY 246 FSVEDTYLL 254  
 Db 658 ---EKSYLI 663  
 RESULT 4  
 500551  
 leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse  
 N:Alternate names: complement-3 receptor alpha chain  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
 C/Accession: S00551; I59078  
 R:Pyteala, R.  
 EMBO J. 7, 1371-1378, 1988  
 A>Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the tr  
 A/Reference number: S00551; MUID:88312584; PMID:3044779  
 A/Accession: S00551  
 A/Molecule type: DNA  
 A/Residues: 1-1153 <PYT>  
 A/Cross-references: UNIPROT:P05555; EMBL:X07640; NID:952982; PIDN:CAA30479.1; PID:952983  
 A/Note: the authors translated the codon CAC for residue 569 as Gln  
 R:Sastre, L.; Roman, J.M.; Teplov, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,  
 Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986  
 A>Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recept  
 A/Reference number: I59078; MUID:86287312; PMID:2942940  
 A/Accession: I59078  
 A/Status: preliminary; translated from GB/EMBL/DDb  
 A/Molecule type: DNA  
 A/Residues: 11-44 <RES>  
 A/Cross-references: GB:M14293; NID:9198993; PIDN:AAA39484.1; PID:9554193  
 C/Genetics:

A/Gene: Mac-1  
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
C/Keywords: cell adhesion; glycoprotein; transmembrane protein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 9.2%; Score 141; DB 2; Length 1153;  
Best Local Similarity 24.2%; Pred. No. 0.0038;  
Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 16;

Qy 18 DLYFIIDKSGSVLHHMNEIYFVEQLAHKFIS---POLRMSFIVSTRGTTLMKLTED-- 72  
Db 150 DIVFLIDSGSIS---NNIDF---QIKMEFVSTWEGKSKSLTFS-----LMQSDSEFR 197  
Qy 73 -----REQIRQGLEELQKLP-----GGDTYMHGSEGRASEQIYYE--NNGQRTASVIATLT 123  
Db 198 IHFTFNDPKRNPSPRSHSPFKQKNGRTKTASGIRKVVRELPHKTNGARENAKILVIT 257  
Qy 124 DGELEHEDLPFYSR--REANRGRDGAIVYCVGVND-FNETQIARIAD-----SKDHVP 174  
Db 258 DGEKFGDPLDYKDIYPEADRA---GVIRYVIGVGNAPKPKPSRRELDTISKAPGEHVQ 314  
Qy 175 VNDGFQALQGIHSLIKKSCIEILAEPSTICAGESEFQVVGNGFRRHARNDVCLSPK 234  
Db 315 V-DNFEALNTIONLOEK---FALIEGTGTSTSFHEWMSGEGF----- 355  
Qy 235 INDVTLNKEKPSVEDTYLCPAFLKEVG---MKALQVSNMDGSLSTSS 283  
Db 356 -SASITSN-----GPILGVSQSPDMAGAFLYTSKDKVFIINTT 393

## RESULT 5

A48569  
antigen Em100 - Eimeria maxima  
C/Species: Eimeria maxima  
C/Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A48569  
R:Paranontes, L.; Hug, D.; Humbelin, M.; Weber, G.  
Mol. Biochem. Parasitol. 57, 171-174, 1993  
A/Title: Sequence of a major Eimeria maxima antigen homologous to the Eimeria tenella m  
A/Reference number: A48569; MUID:93149203; PMID:842611  
A/Accession: A48569  
A/Status: preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1-724 <PAS>  
A/Cross-references: UNIPROT:Q04588; GB:M99058; NID:9158890; PID:9158891  
A/Note: Sequence extracted from NCBI backbone (NCBIN:123776, NCBI:P123777)  
F:45-218/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:238-296/Domain: thrombospondin type 1 repeat homology <THR1>  
F:309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
F:372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
F:433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
F:494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
F:560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 9.0%; Score 139; DB 2; Length 724;  
Best Local Similarity 26.6%; Pred. No. 0.003; Indels 34; Gaps 14;  
Matches 59; Conservative 40; Mismatches 89;

Qy 13 CYGGFDLYFDKSGSV-LHHMNEIYFVEQLAHKF-ISP-QLRMSFIVSTRGTTLMKL 69  
Db 42 CTRLLDVWLVDGSGISITSYKVRFSISNFAQMPSPDVAVGALTFTGSAVTMDL 101  
Qy 70 TEDREQIRQGLEELQKLP--GGDTYMHGSEGRASEQIYYENRQGYR--TASVIATLDG 125  
Db 102 SDSRAQNDLLAAAKKLPAAGSTYTHGLAKA-EBILTFQKGRNAPKMLIVMTDG 160  
Qy 126 ELHEDLPFYSERE-----ANRSDIGAIVYCVGV-KDFNETOLARIA--DSKHV-PP- 174  
Db 161 A-----SSRSQTLIAAEKLNRRGVIIVLVGIGVNSABEGRSINGCDTSTVECPR 212

Qy 175 -VNDGFQALQGIHSLIKKSCIEI---LAAEPSTI--CAGE 209  
Db 213 YLGSNWGVSSQINGIILKAACKDLADAVCSESEYGPCEG 254

RESULT 6  
A45974  
collagen alpha 1(XIV) chain precursor, short form 2 - chicken

N/Alternate names: undulin  
C/Species: Gallus gallus (chicken)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: A45974; S30085; S22916; S17035; S20833  
R:Gercke, D.R.; Foley, J.W.; Caetagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Lin  
J. Biol. Chem. 268, 12177-12184, 1993  
A/Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' regions  
ns.

A/Reference number: A45974; MUID:93280195; PMID:8505337  
A/Accession: A45974  
A/Status: preliminary  
A/Molecule type: mRNA; protein  
A/Residues: 1-1747 <GER>  
A/Cross-references: UNIPROT:P32018  
A/Experimental source: embryo skin  
A/Note: Sequence inconsistent with the nucleotide translation  
A/Note: Sequence extracted from NCBI backbone (NCBIN:133364, NCBI:P133365)  
R:Apfe, S.S.  
submitted to the EMBL Data Library, March 1992  
A/Reference number: S30085

A/Accession: S30085  
A/Molecule type: mRNA  
A/Residues: 1472-1660 <APT>  
A/Cross-references: EMBL:X65122; NID:962871; PID:CAA46238.1; PID:9938175  
R:Trueb, J.; Trueb, B.  
Eur. J. Biochem. 207, 549-557, 1992

A/Title: Type XIV collagen is a variant of undulin.  
A/Reference number: S22916; MUID:92359443; PMID:1339349  
A/Accession: S22916

A/Status: preliminary  
A/Molecule type: mRNA

A/Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>  
R:Gordon, M.K.; Caetagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Rest, M.; Mayne, F

Bur. J. Biochem. 201, 333-338, 1991  
A/Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens v  
A/Reference number: S17035; MUID:92037585; PMID:1935930

A/Accession: S17035  
A/Molecule type: mRNA

A/Residues: 1472-1659 <GOR1>  
A/Accession: S20833

A/Molecule type: protein  
A/Residues: 1551-1570, 1593-1599, 1639-1667 <GOR2>

C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F:40-204/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:236-317/Domain: fibronectin type III repeat homology <FN3A>

F:326-409/Domain: fibronectin type III repeat homology <FN3B>  
F:418-498/Domain: fibronectin type III repeat homology <FN3C>

F:507-591/Domain: fibronectin type III repeat homology <FN3D>  
F:625-707/Domain: fibronectin type III repeat homology <FN3E>

F:716-798/Domain: fibronectin type III repeat homology <FN3F>  
F:806-893/Domain: fibronectin type III repeat homology <FN3G>

F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1111-1353/Domain: non-collagenous NC4 #status predicted <NC4>

F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>  
F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 8.9%; Score 137; DB 2; Length 1747;  
Best Local Similarity 25.4%; Pred. No. 0.014; Indels 26; Gaps 12;  
Matches 63; Conservative 45; Mismatches 114;

Qy 18 DLYFIIDKSGSV-LHHMNEIYFVEQL--AHKISFQ-LRMSFIVSTRGTTLMKL--E 71  
Db 926 DIVFLVDGWSISDIDDFNKISFLYSTVGALDYGIDGQVALIQSDDDRTFKNAYK 985  
Qy 72 DREQIRQGLEELQKLPVGGDTYMHGSEGRASEQIYYENRQGYR--SVIATLDGELHE 129

Db 966 TKEITLLEIQQI--AYKGGNTKTGAIGARE-VLFTSEAGRRKCIPIKVLVITIDGRSD 1042  
Qy 130 DLFFYSSEANRSRL-GAIYVCVGDENFETOLARIAD--SKDHVPVNGFOALQGI 186  
Db 1043 DV-----NKVRREMLDGFSPFAIVADADSELVNISKSSERVAFVDD-FDAFTKIE 1096  
Qy 187 HSILKSCIRILAAEPSTICAGESFQVYVNGNGFPAHANVDRVLCSFKINDSVTLNEKPF 246  
Db 1097 DELTIFVCETASATCLPLVKQDNFA-----GFGMGEMFGLVEKFSALDGVSNPEPTF 1150  
Qy 247 SVEDTYLL 254  
Db 1151 NVYPCYRL 1158

## RESULT 7

A45638  
Immunodominant microneme protein Exp100 - Eimeria tenella  
C/Species: Eimeria tenella  
C/Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A45638  
R/Tonley, F.M.; Clarke, L.E.; Kawazoe, U.; Djikema, R.; Kok, J.J.  
Mol. Biochem. Parasitol. 49, 277-288, 1991  
A/Title: Sequence of the gene encoding an immunodominant microneme protein of Eimeria tenella  
A/Reference number: A45638; MUID:92131064; PMID:1775171  
A/Accession: A45638  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-712 <TM>  
A/Cross-references: UNIPROT:O43981; GB:AFO32905; GB:M73495; NID:92707732; PIDN:AAD03350.  
F/Note: Sequence extracted from NCBI backbone (NCBI:77752, NCBI:77756)  
F/48-218/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/238-296/Domain: thrombospondin type 1 repeat homology <THR1>  
F/309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
F/372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
F/433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
F/494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
F/560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 8.8%; Score 134.5; DB 2; Length 712;  
Best Local Similarity 24.6%; Pred. No. 0.0068;  
Matches 55; Conservative 40; Mismatches 88; Indels 41; Gaps 11;  
Qy 13 CYGGFDLYFLDKSGV-LHHMNEIYFVEOLAHKF-ISPQ-LRMSFVFSRGTLMKL 69  
Db 45 CTSLDLVLLVDSGSIGTSNFRKVRQPIEDPNSMPSPEVNRGLITFT----- 96  
Qy 70 TEDREQIRQGLEBLQKVLPG-----GDTVMHGFERASEQIYYENRQYR-- 114  
Db 97 ---RSKVRWNLSDPKATVPSLAISAARLSYSTGVTHYGLQDA-KKLVDTNAGANN 152  
Qy 115 TASVYIALTDELHEDLFFYSERANRSRDIGAIYVCV-KDNETOLARIADSKHVF 173  
Db 153 VPKVLVLTWTDAA--SNLPSQTRSSAAALRDGAIVVVGSGVNSSECRSIAGCSTNC 210  
Qy 174 P-----VNDGFQALQGIHSLKSCIRILAAEPSTI--CAGE 209  
Db 211 PRLQSNMNSNTQVNGIILKACQDLADAVCSEMSYGPVGE 254

## RESULT 8

C2MS  
Classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2 component precursor - mc  
N/Alternate names: C3 convertase; C5 convertase; complement C2  
C/Species: Mus musculus (house mouse)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: A38876; B36593; I54429  
R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
Submitted to GenBank, January 1991  
A/Reference number: A38875  
A/Accession: A38876  
A/Molecule type: DNA

A/Residues: 1-760 <IS2>  
A/Cross-references: UNIPROT:P21180; GB:M57891; GB:J05661; NID:9192436; PIDN:AAA63294.1; R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
J. Biol. Chem. 265, 19040-19046, 1990  
A/Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different met  
A/Reference number: A36593; MUID:91035430; PMID:2229060  
A/Accession: B36593

A/Molecule type: mRNA  
A/Residues: 1-760 <ISH>  
A/Cross-references: EMBL:M57891; NID:9192436; PIDN:AAA63294.1; PID:9192437  
R/Falus, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, D.; Whitehead, A.S.; Colten, H.R.  
Immunogenetics 25, 230-268, 1987  
A/Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.  
A/Reference number: I54429; MUID:87192938; PMID:2883115  
A/Accession: I54429

A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 660-677, R', 679, 681-723, G', 725 <RBS>  
A/Cross-references: GB:M16271; NID:9199289; PIDN:AAA39562.1; PID:9199290  
C/Genetics:

A/Introns: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 414/1; 492/3; 531/3  
C/Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C2a,  
C/Function:  
A/Description: cleaves complement C3 and complement C5 alpha chains  
A/Pathway: complement classical pathway  
C/Suprafamily: complement B/C2; complement factor H repeat homology; trypsin homology; VC  
C/Keywords: alternative splicing; complement classical pathway; duplication; glycoprotein  
F/1-18/Domain: signal sequence #status predicted <SIG>  
F/19-250/Product: complement C2b fragment #status predicted <C2b>  
F/251-250/Product: complement C2a fragment long form #status predicted <C2a>  
F/251-605, 613-760/Product: complement C2a fragment short form #status predicted <C2s>  
F/259-449/Domain: von Willebrand factor type A repeat homology <VFA>  
F/478-747/Domain: trypsin homology #status atypical <TRY>  
F/2-62, 49-89, 94-136, 122-149, 156-197, 182-210, 470-590, 499-515, 593-609, 647-674, 685-715/Dia  
F/27, 117, 297, 340, 474, 478, 663/Binding site: carbohydrate (Asn) (covalent) #status predict  
F/514, 570, 689/Active site: His, Asp, Ser #status predicted

Query Match 8.5%; Score 131; DB 1; Length 760;  
Best Local Similarity 23.2%; Pred. No. 0.014;  
Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;

Qy 15 GGFDFLYFLDKSGV-LHHMNEIY-FVEOLAHKFIPOLR--MSFVFSRGTLMKL 71  
Db 258 GHLNLYLLLDASQSVTEKDPDIFKSNALMERIFSEVNTVAIITFASOPTIMILS 317  
Qy 72 DREQIRQGLEBLQKVLPGCDTVMHGFERASEQIYYE-----NRQYRTAS- 117  
Db 318 ERSG-----DVEVITSLDSASVYKHENATGANTYEVLLRVYSMOTQDRIGMETSAA 371  
Qy 118 ----VIALTDELHEDLFFYSERANRSRDIGAI-----VYCVG-----KD 156  
Db 372 KEIRHTIILLTDGK--SNMGSPPKAVTRIRBELSLIQNRDYLIDYIAGVGKLDVWKE 429  
Qy 157 FNERQIARIADSKOHVPVNDGFOALQGIHSLKSCIRILAAEPSTICAGESFQVYV 216  
Db 430 LNE--LSSKQDGEHATILQDA-KALQOIFENHMDVSKL-----TDTIG-----V 472  
Qy 217 GNGFRHARANDRV--LCSFKINDSVTLNEKPSVEDTYLLCPAFLKEVGK--AAQVS 272  
Db 473 GNSMANSDDERTFWQYTFPKSKETQGS--LISDQVLTAAHCFDIDGMDHHLRVN 530  
Qy 273 MND-----GLSFISSSVYI 286  
Db 531 VGPDSQHKRFLVEDVYI 549

## RESULT 9

S31212  
collagen alpha 1(XIV) chain precursor, short form - chicken  
C/Species: Gallus gallus (chicken)

C>Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 15-Sep-2003  
C/Accession: S31212  
R/Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.  
Eur. J. Biochem. 212, 483-490, 1993  
A>Title: Complete primary structure of chicken collagen XIV.  
A/Reference number: S31211; MUID:93185668; PMID:8444186  
A/Accession: S31212  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-1857 <WAE>  
A/Cross-references: EMBL:X70792; NID:9288874; PIDN:CA50064.1; PID:9288875  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C/Genetics:  
A/Gene: Col14A1  
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F/1-28/Domain: signal sequence #status predicted <SIG>  
F/129-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>  
F/129-110/Domain: fibronectin type III repeat homology <FN3A>  
F/156-330/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/352-433/Domain: fibronectin type III repeat homology <FN3B>  
F/442-525/Domain: fibronectin type III repeat homology <FN3C>  
F/534-614/Domain: fibronectin type III repeat homology <FN3D>  
F/623-707/Domain: fibronectin type III repeat homology <FN3E>  
F/741-823/Domain: fibronectin type III repeat homology <FN3F>  
F/832-914/Domain: fibronectin type III repeat homology <FN3G>  
F/922-1009/Domain: fibronectin type III repeat homology <FN3H>  
F/1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1857;  
Best Local Similarity 25.0%; Pred. No. 0.055;  
Matches 62; Conservative 44; Mismatches 116; Indels 26; Gaps 12;

QY 18 DLVFIIDKSGSV-LHHMNEIYFVEQL-AHKFISPO-LRMSFIVSTRGTTLMKLT--E 71  
DB 1042 DLVFLVDGWSMSIDDDNFKKISFLYSTVGALDKIGDGTQVAIIQFSDPRTFKLNAYK 1101  
QY 72 DREQIRGLELEOKVLPFGDPTVMHEGFERASEQIYENRGYRTA--SVIIALTGSELHE 129  
DB 1102 TKETLLLEAIQGI--AVKGGTKTKGKAIKHARE-VLFTGEAGMKRGIPKVLVITDGRSQD 1158  
QY 130 DLFFYSERANRSDI-GAIVYCVGVKDFMETOLAIAD--SKDHYFPVNDGQALOGIT 186  
DB 1159 DV-----NKVSRMQLDGFSPFAIGVADADYSELVNGSKPSRHVFFVND-PDAFTKIE 1212  
QY 187 HSILKSCIEIIAEPSTICAGESFQVNVGNGFERHARNDVYLCSEFKINDSVTLNEKPF 246  
DB 1213 DELITPVCETASATCPVFKDGLKA-----GFQMMEMGLVKEKFSALDGVSMERCTF 1266  
QY 247 SVEDTYLL 254  
DB 1267 NVYPCYRL 1274

## RESULT 10

S78476 collagen alpha 1(XIV) chain precursor, long form - chicken  
C/Species: Gallus gallus (chicken)  
C/Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 09-Jul-2004  
C/Accession: S78476; S31211  
R/Trueb, B.  
submitted to the EMBL Data Library, January 1993  
A/Reference number: S78476  
A/Accession: S78476  
A/Molecule type: mRNA  
A/Residues: 1-1888 <TRU>  
A/Cross-references: UNIPROT:P32018; EMBL:X70793; NID:9288872; PIDN:CA50064.1; PID:9288875  
R/Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.  
Eur. J. Biochem. 212, 483-490, 1993  
A>Title: Complete primary structure of chicken collagen XIV.  
A/Reference number: S31211; MUID:93185668; PMID:8444186  
A/Accession: S31211  
A/Status: preliminary  
A/Molecule type: mRNA

A/Residues: 1-416;1460-1811,1843-1888 <WAE>  
A/Cross-references: EMBL:X70793  
C/Genetics:  
A/Gene: Col14A1  
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F/1-28/Domain: signal sequence #status predicted <SIG>  
F/129-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>  
F/129-110/Domain: fibronectin type III repeat homology <FN3A>  
F/156-330/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/352-433/Domain: fibronectin type III repeat homology <FN3B>  
F/442-525/Domain: fibronectin type III repeat homology <FN3C>  
F/534-614/Domain: fibronectin type III repeat homology <FN3D>  
F/623-707/Domain: fibronectin type III repeat homology <FN3E>  
F/741-823/Domain: fibronectin type III repeat homology <FN3F>  
F/832-914/Domain: fibronectin type III repeat homology <FN3G>  
F/922-1009/Domain: fibronectin type III repeat homology <FN3H>  
F/1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1888;  
Best Local Similarity 25.0%; Pred. No. 0.057;  
Matches 62; Conservative 44; Mismatches 116; Indels 26; Gaps 12;

QY 18 DLVFIIDKSGSV-LHHMNEIYFVEQL-AHKFISPO-LRMSFIVSTRGTTLMKLT--E 71  
DB 1042 DLVFLVDGWSMSIDDDNFKKISFLYSTVGALDKIGDGTQVAIIQFSDPRTFKLNAYK 1101  
QY 72 DREQIRGLELEOKVLPFGDPTVMHEGFERASEQIYENRGYRTA--SVIIALTGSELHE 129  
DB 1102 TKETLLLEAIQGI--AVKGGTKTKGKAIKHARE-VLFTGEAGMKRGIPKVLVITDGRSQD 1158  
QY 130 DLFFYSERANRSDI-GAIVYCVGVKDFMETOLAIAD--SKDHYFPVNDGQALOGIT 186  
DB 1159 DV-----NKVSRMQLDGFSPFAIGVADADYSELVNGSKPSRHVFFVND-PDAFTKIE 1212  
QY 187 HSILKSCIEIIAEPSTICAGESFQVNVGNGFERHARNDVYLCSEFKINDSVTLNEKPF 246  
DB 1213 DELITPVCETASATCPVFKDGLKA-----GFQMMEMGLVKEKFSALDGVSMERCTF 1266  
QY 247 SVEDTYLL 254  
DB 1267 NVYPCYRL 1274

## RESULT 11

S42373 hypothetical protein T20G5.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-May-2004  
C/Accession: S42373  
R/Smith, A.  
submitted to the EMBL Data Library, March 1994  
A/Reference number: S42368  
A/Accession: S42373  
A/Molecule type: DNA  
A/Residues: 1-3051 <SMT>  
A/Cross-references: EMBL:Z30423; NID:9458479; PID:9458485  
C/Genetics:  
A/Insertions: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/1; 952/2; 1041/1; 1146/1;  
F/512-679/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/754-793/Domain: fibronectin type II repeat homology <2FI>  
F/1201-1244/Domain: EGF homology <EGF>

Query Match 8.3%; Score 128; DB 2; Length 3051;  
Best Local Similarity 27.1%; Pred. No. 0.16;  
Matches 64; Conservative 31; Mismatches 75; Indels 66; Gaps 13;

QY 13 CYGGE-----DLVFIIDKSGSVLHH--MNEIYFVEQLA 44  
DB 483 CYGGEFVNVSSNANLPGRVCTVQTTCPKQKTDIVFLVDGSGSGSYFKNIEVLAIFVREFEV 542  
QY 45 HKFI--SPOLRMSFIVSTRGTTLMKLT--DREQIRGLELEOKVLPG-----GDTVMH- 95  
DB 543 ELFEIGSKTRVGLIQYSDQIRHFFLDQYGRDLSLKGISFTQ-YLTGLTRGAALQHM 601





A:Reference number: A19188; MUID:83204002; PMID:6342610  
A:Contents: the final paper in a series documenting the sequence, glycosylation site, at  
A:Accession: A19188  
A:Molecule type: protein  
A:Residues: 260-286, 'T', 298-764 <CHR>  
R:Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A:Title: Molecular cloning and characterization of the gene coding for human complement  
A:Reference number: A19947; MUID:83272641; PMID:6308626  
A:Accession: A19947  
A:Molecule type: DNA  
A:Residues: 346-764 <CAM>  
A:Cross-references: GB:J00125  
A:Accession: B19947  
A:Molecule type: mRNA  
A:Residues: 339-509 <CA1>  
A:Cross-references: GB:J00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
R:Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A:Title: Cell-specific expression of the human complement protein factor B gene: evidenc  
A:Reference number: A25971; MUID:87102880; PMID:3643061  
A:Accession: B25971  
A:Molecule type: DNA  
A:Residues: 1-99 <WU>  
A:Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A:Title: The principal site of glycation of human complement factor B.  
A:Reference number: S14339; MUID:91174758; PMID:2006911  
A:Accession: S14339  
A:Molecule type: protein  
A:Residues: 270-339 <NIE>  
A:Note: binding site for carbohydrate to lysine under artificial conditions  
R:Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A:Title: Internal homologues of the Ba fragment from human complement component factor B  
A:Reference number: A44628; MUID:84158524; PMID:6323161  
A:Accession: A44628  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 16-225, 'P', 227-259 <MOR>  
R:Schwable, W.; Luttig, B.; Sokolowski, T.; Escallier, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A:Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A:Reference number: I54409; MUID:94041399; PMID:8225386  
A:Accession: I54409  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: GB:S67310; NID:g452937; PIDN:AMD13989.1; PID:g4261689  
R:Horluchi, T.; Kilm, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conv  
A:Reference number: I57824; MUID:94067177; PMID:8247029  
A:Accession: I57824  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-31, 'O', 33-764 <RE2>  
A:Cross-references: GB:U5702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C:Comment: 292-Cys has a free sulphydryl.  
C:Genetics:  
A:Gene: GDB:BF  
A:Cross-references: GDB:119726; OMIM:138470  
A:Map position: 6p21.3-6p21.3  
A:Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 65  
A:Note: the 11st of introns may be incomplete  
C:Complex: complement factor B initially forms an inactive complex with complement facto  
ment factor C3b forming active C3/C5 convertase, Ba is released  
C:Function:  
A:Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha c  
A:Pathway: complement alternate pathway

C:Superfamily: complement B/C3; complement factor H repeat homology; trypsin homology; v  
C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydrol  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <MAT>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <FH1>  
F:103-158/Domain: complement factor H repeat homology <FH2>  
F:165-218/Domain: complement factor H repeat homology <FH3>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
F:268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76, 62-98, 103-145, 131-158, 165-205, 191-218, 478-596, 511-527, 599-615, 656-682, 695-725/D  
F:122, 142, 285, 378/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526, 576, 699/Active site: His, Asp, Ser #status experimental  
Query Match 7.4%; Score 113.5; DB 1; Length 764;  
Best Local Similarity 19.6%; Pred. No. 0.37;  
Matches 66; Conservative 63; Mismatches 109; Indels 99; Gaps 17;  
QY 2 OGRRRDG-GPA-----CYGFPDYFLDKSGSV-----LHNMNRYFVEQLAH 45  
DB 243 EGVDAEDGHGPRGQKRIYLDPSGSMITYLVDSDISGASFTGAKKCLVILKRVAS 302  
QY 46 KRISPOLRMSFVFSRGTTLTKLT-----DREQIRQGLEEL-----QKVLPGSDTYMHG 97  
DB 303 YGVKP--RYGLVYATYATPKIYVWSEADSSNADWVYKQLEINVEDHKKSGTNT----- 355  
QY 98 FEASRQIYENR-----QGY-RTASVITLTDG-----ELHBDLPFYSE 136  
DB 356 -KRALDQVYVSMWSPPDVPPEGNNRTHVILMTDGLHNMGDPITVIDEIRLLIYKGD 414  
QY 137 REANRSRDGAIIYCVG--VKDFNETOLARIADSKOHVPEVNDFOALQGIHSLIKKSC 194  
DB 415 RKNRPREDYLDVYFVGPLVQVNIINALSKQNEGVFVVD--MENTLBDVFPQWIDES- 472  
QY 195 IEILAAPEPTTCAGSEFQVYVWGRNGRHHANDRVLCSEKINDSVTLNKPSEVEDTYLL 254  
DB 473 -----QSLSLC-----GMVWEHRKGTD-----YHKPQWAKISV-- 501  
QY 255 CPAPILKEVGMKALQVNMNDGLSFSSSVIITTCG 291  
DB 502 ----IRPSKHESYCMG-----AVVSEYFVLTAHC 527  
RESULT 15  
A54849  
collagen alpha 1(VII) chain precursor - human  
N:Alternate names: procollagen alpha 1(VII) chain  
C:Species: Homo sapiens (man)  
C:Date: 04-Nov-1994 #sequence revision 04-Nov-1994 #text change 09-Jul-2004  
A:Accession: A54849; PH0844; S16316; I56328; A30296; I84686  
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.  
J. Biol. Chem. 269, 20256-20262, 1994  
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)  
A:Reference number: A54849; MUID:94327588; PMID:8051117  
A:Accession: A54849  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2944 <CHR>  
A:Cross-references: UNIPROT:Q02388; GB:L02870; NID:g987124; PIDN:AAA75438.1; PID:g987125  
R:Tanaka, T.; Takahashi, K.; Futukawa, F.; Imanura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A:Title: Molecular cloning and characterization of type VII collagen cDNA.  
A:Reference number: PH0844; MUID:92231902; PMID:1567409  
A:Accession: PH0844  
A:Molecule type: mRNA  
A:Residues: 'EPR', 340-475, 'RALSTASHSTLCRRATRRHPCNCGSHWTRACPCNRPASHRAARAG', 524-528, 'C',  
A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:g453698; PIDN:BA020853.1; PID:g453699  
A:Experimental source: keratinocyte  
A:Note: the authors translated the codon ACC for residue 394 and 397 as Tyr  
R:Parente, M.G.; Chung, L.C.; Rymaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991



A/Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A/Reference number: S16316; MUID:91334380; PMID:1871109  
A/Accession: S16316  
A/Molecule type: mRNA  
A/Residues: 815-892 'E', 894-1439 <PAR>  
A/Cross-references: GB:M55158; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
A/Experimental source: Keratinocyte  
R/Gammon, W.R.; Abernethy, M.L.; Pedilla, K.M.; Priyaanah, P.S.; Cook, M.E.; Wright, J.;  
J. Invest. Dermatol. 99, 691-696, 1992  
A/Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot  
A/Reference number: S16328; MUID:93107742; PMID:1465284  
A/Accession: S16328  
A/Status: translated from GB/EMBL/DBJ  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
A/Cross-references: GB:S51236; NID:G262308; PIDN:AA24637.1; PID:G262309  
R/Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Gnanville, R.W.; Burgess, R.E.  
J. Biol. Chem. 264, 3822-3826, 1989  
A/Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
A/Reference number: A30296; MUID:89139437; PMID:2537292  
A/Accession: A30296  
A/Molecule type: Protein  
A/Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, '  
A/Note: two reported peptides cannot be reliably located  
R/Greenpan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A/Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
A/Reference number: I48103; MUID:93271985; PMID:8499916  
A/Accession: I48103  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 2395-2871, 'S', 2873-2944 <RE2>  
A/Cross-references: GB:L06862; NID:G388713; PIDN:AAA89196.1; PID:G388714  
R/Christiano, A.M.; Rymaszewski, M.; Uitto, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A/Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser sub  
A/Reference number: A55255; MUID:94224777; PMID:8170945  
A/Contents: annotation  
A/Comments: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C/Genetics:  
A/Gene: GDB:COL7A1; EBR1; EBD1; EB  
A/Cross-references: GDB:128750; OMIM:120120  
A/Map position: 3p21.3-3p21.3  
A/Note: defects in this gene can result in dominant and recessive dystrophic epidermolys  
A/Note: there are 118 introns  
C/Complex: type VII collagen is probably a homotrimer  
C/Function:  
A/Description: structural component of extracellular polymer associated with anchoring f  
C/Keywords: coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli  
F/1-16/Domain: signal sequence #status predicted <SIG>  
F/17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F/17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F/36-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/231-318/Domain: fibronectin type III repeat homology <FN1>  
F/337-413/Domain: fibronectin type III repeat homology <FN2>  
F/414-502/Domain: fibronectin type III repeat homology <FN3>  
F/508-593/Domain: fibronectin type III repeat homology <FN4>  
F/598-683/Domain: fibronectin type III repeat homology <FN5>  
F/686-771/Domain: fibronectin type III repeat homology <FN6>  
F/776-862/Domain: fibronectin type III repeat homology <FN7>  
F/864-952/Domain: fibronectin type III repeat homology <FN8>  
F/954-1045/Domain: fibronectin type III repeat homology <FN9>  
F/1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
F/1170-1172/Domain: cell attachment (R-G-D) motif  
F/1189-1253/Domain: cytochrome/retinol-rich  
F/1254-2783/Region: interrupted helical  
F/1334-1336/Region: cell attachment (R-G-D) motif  
F/2008-2010/Region: cell attachment (R-G-D) motif  
F/2553-2555/Region: cell attachment (R-G-D) motif  
F/2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F/2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
F/337, 786, 1109/Binding site: carbohydrate (Aan) (covalent) #status predicted

F/2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper  
F/2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental  
F/2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F/2634,2802,2804/Dissulfide bonds: interchain #status predicted

Query Match 7.3%; Score 111.5; DB 2; Length 2944;  
Best Local Similarity 24.1%; Pred. No. 3.2;  
Matches 52; Conservative 42; Mismatches 85; Indels 37; Gaps 11;  
QY 18 DLYETILKSGSV-LHNNNEIYFVEOLAHNF---ISPO-LRMSIVSTRET---LMKL 69  
DB 38 DIVFLDSSSIGRSNFRVRSFLEGLVLPFGSAAGVRFATVQYSDPRTFGLDAL 97  
QY 70 TEDREQIRQGLEBLQKVLPGSDTYMHGFEFASQIYENRQGYRTASVIALTDGELHE 129  
DB 98 GSGGDVIR-AIRELS--YKGNRTTGAALLHVAHVLPLQARPGVKCLIRDGK-SQ 153  
QY 130 DLFYSREANRSDLAIIYCVGKOPNETOLARIID--SKHVFVNDGFQALQGIH 187  
DB 154 DLV--DTAAQRLEKGGQVKLFAVGIKNADPEELKRVASQPTSDFFPVND-FSILRTLLP 209  
QY 188 SILKSGI-----EILAEPSST 204  
DB 210 LVSRRVCTTAGGVPTRPDDSTISAPRDLVLSPESS 245

Search completed: June 13, 2005, 20:06:35  
Job time : 23.1647 secs

**THIS PAGE BLANK (USPTO)**



```

XX 13-JUN-2002.
XX
XX 03-OCT-2001; 2001WO-US030941.
XX
XX 05-DEC-2000; 2000US-0251481P.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
XX Young JAT, Bradley KA, Collier RJ, Mogridge JS;
XX
XX WPI: 2002-713235/77.
XX
XX N-PSDB; ABV73881.
XX
XX Novel isolated polypeptide useful for identifying agent that prevents or
XX reduces effect of anthrax toxin on host cell, for treating human or non-
XX human animal suffering from anthrax.
XX
XX Claim 1; Page 29-30; 45pp; English.
XX
XX The present sequence is the protein sequence of a human surface-bound
XX anthrax toxin receptor (ATR), as predicted from an isolated cDNA clone.
XX Anthrax toxin protective antigen (PA) binds to the ATR at a von
XX Willebrand factor A domain located in the extracellular domain of ATR.
XX The invention provides ATR polypeptides and polynucleotides, vectors,
XX host cells, and transgenic and knock-out animals. It also provides
XX methods for identifying molecules that bind the ATR and which reduce the
XX toxicity of anthrax toxin. A claimed method for treating anthrax in a
XX human or animal involves administering an agent that inhibits binding
XX between PA and ATR at a level effective to reduce the severity of
XX anthrax. Suitable agents include ATR or a PA-binding fragment of ATR, a
XX PA-binding polypeptide at least 80% identical to these, a fusion protein,
XX a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a
XX nucleic acid
XX
XX Sequence 368 AA:
XX
XX Query Match 100.0%; Score 1526; DB 5; Length 368;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-155;
XX Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 QGGRBDGPGACGCGFDLYFLDSSGSLHHMNEYVVEQLAKFTSPQIRMFYEST 60
XX |
XX 28 QGGRBDGPGACGCGFDLYFLDSSGSLHHMNEYVVEQLAKFTSPQIRMFYEST 87
XX
XX 61 RGTLMKLTERRBOIRQGLBELQKVLPGCDTYMEHGFERRASEQIYYENRQGYRTASV11A 120
XX |
XX 88 RGTLMKLTERRBOIRQGLBELQKVLPGCDTYMEHGFERRASEQIYYENRQGYRTASV11A 147
XX
XX 121 LTDEGLHEDLFFYSERREANRSDIGAIYCVGVDFNETQIARIADSKDHVPVNDGFOA 180
XX
XX 148 LTDEGLHEDLFFYSERREANRSDIGAIYCVGVDFNETQIARIADSKDHVPVNDGFOA 207
XX
XX 181 LOGIITHLTKKSCIEIIAAEPSTTCAGSFGVVVVRNGNFRARAVNDRLCSFKINDSVTL 240
XX
XX 208 LOGIITHLTKKSCIEIIAAEPSTTCAGSFGVVVVRNGNFRARAVNDRLCSFKINDSVTL 267
XX
XX 241 NEKFPVSVDYTLCPAPILKEVGKMAALQVMNDGLSISSSVIITTHCSDG 293
XX |
XX 268 NEKFPVSVDYTLCPAPILKEVGKMAALQVMNDGLSISSSVIITTHCSDG 320
XX
XX
XX RESULT 2
XX AAE01439 standard; protein; 403 AA.
XX
XX AAE01439;
XX
XX 17-JUN-2001 (first entry)
XX
XX Human gene 4 encoded secreted protein HMLFR02, SEQ ID NO:94.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;

```

```

XX foetal abnormality; developmental abnormality; hematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder;
XX endocrine disorder; infection; wound healing; vulnery; cell culture;
XX chemotaxis; food additive; gene therapy; binding partner identification;
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..27
XX Protein /label= signal_peptide
XX 28..403
XX /note= "Mature human secreted protein"
XX
XX W0200134626-A1.
XX
XX 17-MAY-2001.
XX
XX 01-NOV-2000; 2000WO-US030045.
XX
XX 05-NOV-1999; 99US-0163581P.
XX
XX 30-JUN-2000; 2000US-0215133P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruden SM, Komatsuolis GA, Moore PA, Birse CE, Nt J;
XX
XX WPI: 2001-308778/32.
XX
XX N-PSDB; AAD05303.
XX
XX New nucleic acid molecules encoding 28 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and
XX used as food additives or preservatives.
XX
XX Claim 11; Page 485-486; 562pp; English.
XX
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted
XX protein genes, and AAE01436-AAE01513 represent the proteins they encode.
XX AAE01514-AAE01544 represent human secreted protein fragments or variants.
XX The genes and their secreted proteins are useful for preventing, treating
XX or ameliorating medical conditions, e.g., by protein or gene therapy.
XX Pathological conditions can be diagnosed by determining the amount of the
XX new protein in a sample or by determining the presence of mutations in
XX the new genes. Specific uses are described for each of the 28 genes,
XX based on the tissues in which they are most highly expressed, and include
XX developing products for the diagnosis or treatment of proliferative
XX disorders, cancer, tumours, foetal and developmental abnormalities,
XX hematopoietic disorders, diseases of the immune system, AIDS, autoimmune
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
XX angiogenic disorders, kidney disorders, gastrointestinal disorders,
XX pregnancy-related disorders, endocrine disorders, and infections. The
XX protein can also be used to aid wound healing and epithelial cell
XX proliferation, to prevent skin aging due to sunburn, to maintain organs
XX before transplantation, for supporting cell culture of primary tissues,
XX to regenerate tissues, to identify their cognate ligands or binding
XX partners, and in chemotaxis, and can be used as a food additive or
XX preservative to modify storage properties. Antibodies specific for a
XX protein of the invention can be used in alleviating symptoms associated
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
XX present sequence represents a human secreted protein of the invention
XX
XX Sequence 403 AA:
XX
XX Query Match 100.0%; Score 1526; DB 4; Length 403;

```

Best Local Similarity 100.0%; Pred. No. 1.8e-155; Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QGGRREDGPACGPGFDLYFLDKSGSVLHHMNEIYFVQOLAHKFISPOLRMSFIYFST 60  
Dd 28 QGGRREDGPACGPGFDLYFLDKSGSVLHHMNEIYFVQOLAHKFISPOLRMSFIYFST 87  
QY 61 RGTTLMLKLTEDREQIRQGLEBLQVLPFGDPTVMEHGFERSAQIYYENRQGYRTASVITIA 120  
Dd 88 RGTTLMLKLTEDREQIRQGLEBLQVLPFGDPTVMEHGFERSAQIYYENRQGYRTASVITIA 147  
QY 121 LTDELHEDLPFYSEBRANSRDLGAIYVCVGDVFNFTOLARLADSKDHVPVNDGFQA 180  
Dd 148 LTDELHEDLPFYSEBRANSRDLGAIYVCVGDVFNFTOLARLADSKDHVPVNDGFQA 207  
QY 181 LQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRGNGFRRARVNDVLCFSKINDSVTL 240  
Dd 208 LQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRGNGFRRARVNDVLCFSKINDSVTL 267  
QY 241 NEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSGD 293  
Dd 268 NEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSGD 320

RESULT 3  
ABG63874 ID ABG63874 standard; protein; 403 AA.

XX AC ABG63874;

XX DT 27-AUG-2002 (first entry)

XX DE Human albumin fusion protein #549.

XX KM Albumin fusion protein; therapeutic protein X; human albumin; HA;  
KM human serum albumin; HSA; cancer; reproductive disorder;  
KM digestive disorder; immune disorder; endocrine disorder;  
KM haematopoietic disorder; neural disorder; connective disorder;  
KM cytoskeletal; antiinfectivity; antiinflammatory; anticancer;  
KM immunomodulator; anti-HIV; antidiabetic; haemostatic; noctropic;  
KM neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
KM osteopathic; antichratic.

XX OS Homo sapiens.  
XX OS Synthetic.

XX PN MO200177137-A1.

XX PD 18-OCT-2001.

XX PF 12-APR-2001; 2001WO-US011988.

XX PR 12-APR-2000; 2000US-0229358P.

XX PR 25-APR-2000; 2000US-0199384P.

XX PR 21-DEC-2000; 2000US-0256931P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Haseltine WA;

XX DR WPI; 2002-010886/01.

XX PT New fusion protein for treating disease e.g. diabetes comprises an

XX XX Claim 1, Page 874-875, 2102pp; English.

XX CC The present invention relates to albumin fusion proteins comprising a  
XX CC therapeutic protein X and human albumin (HA), also known as human serum  
XX CC albumin, HSA). The proteins are useful for treating a disease or disorder  
XX CC that may be modulated by therapeutic protein X. The albumin extends the  
XX CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
XX CC activity. The protein is useful for treating and diagnosing disorders

CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
CC disease, ulcerative colitis), immune disorders (e.g. acquired  
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention

XX S0 Sequence 403 AA;  
Query Match 100.0%; Score 1526; DB 5; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.8e-155; Indels 0; Gaps 0;  
Matches 293; Conservative 0; Mismatches 0;

QY 1 QGGRREDGPACGPGFDLYFLDKSGSVLHHMNEIYFVQOLAHKFISPOLRMSFIYFST 60  
Dd 28 QGGRREDGPACGPGFDLYFLDKSGSVLHHMNEIYFVQOLAHKFISPOLRMSFIYFST 87  
QY 61 RGTTLMLKLTEDREQIRQGLEBLQVLPFGDPTVMEHGFERSAQIYYENRQGYRTASVITIA 120  
Dd 88 RGTTLMLKLTEDREQIRQGLEBLQVLPFGDPTVMEHGFERSAQIYYENRQGYRTASVITIA 147  
QY 121 LTDELHEDLPFYSEBRANSRDLGAIYVCVGDVFNFTOLARLADSKDHVPVNDGFQA 180  
Dd 148 LTDELHEDLPFYSEBRANSRDLGAIYVCVGDVFNFTOLARLADSKDHVPVNDGFQA 207  
QY 181 LQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRGNGFRRARVNDVLCFSKINDSVTL 240  
Dd 208 LQGIHSLKSKSCIEILAAEPSTICAGESFOVVVRGNGFRRARVNDVLCFSKINDSVTL 267  
QY 241 NEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSGD 293  
Dd 268 NEKPFVSDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSGD 320

RESULT 4  
ADL77139 ID ADL77139 standard; protein; 403 AA.

XX AC ADL77139;

XX DT 20-MAY-2004 (first entry)

XX DE Albumin fusion protein related therapeutic protein X, SEQ ID No 621.

XX KM Albumin fusion protein; cytostatic; antianaemic; antichratic;  
KM antiaetmatic; anti-HIV; immunosuppressive; antiinflammatory;  
KM antipsoaratic; antibacterial; osteopathic; dermatological; antigout;  
KM immunomodulator; antiarhythmic; cardiac; noctropic; antilipemic;  
KM nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;  
KM antidiabetic; anabolic; hypertensive; vulnery; gene therapy; cancer;  
KM reproductive system disorder; therapeutic protein.

XX OS Undefined.

XX PN US2004010134-A1.

XX PD 15-JAN-2004.

XX PF 12-APR-2001; 2001US-00833245.

XX PR 12-APR-2000; 2000US-0229358P.

XX PR 25-APR-2000; 2000US-0199384P.

XX PR 21-DEC-2000; 2000US-0256931P.

XX PA (ROSE/) ROSEN C A.

XX PI (HASE/) HASELTINE W A.

XX DR Rosen CA, Haseltine WA;

XX XX WPI; 2004-090519/09.  
XX PT New albumin fusion proteins, useful for diagnosing, creating, preventing

PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,  
PT asthma, inflammatory bowel disease or Alzheimer's disease.  
PS Disclousure; SEQ ID NO 621; 2799p; English.  
XX  
CC The invention relates to a novel albumin fusion protein. The invention  
CC further relates to: a composition comprising the albumin fusion protein  
CC and a pharmaceutical carrier; a kit comprising the composition of the  
CC albumin fusion protein formula; a method of treating a disease or  
CC disorder in a patient comprising the step of administering the albumin  
CC fusion protein; a method of treating a patient with a disease or disorder  
CC that is modulated by therapeutic protein: X, or its fragment or variant;  
CC a method of extending the shelf life of Therapeutic protein: X, or its  
CC fragment or variant; a nucleic acid molecule comprising a polynucleotide  
CC sequence encoding the albumin fusion protein; a vector comprising the  
CC nucleic acid molecule of the albumin fusion protein; and a host cell  
CC comprising the nucleic acid molecule of the albumin fusion protein. The  
CC albumin fusion protein and its compositions have the following  
CC activities: cytostatic, antineoplastic, antitumor, antidiabetic, anti-  
CC HIV, immunosuppressive, antiinflammatory, antiproliferative, antibacterial,  
CC osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic,  
CC neuroprotective, antiparkinsonian, tranquilizer, uropathic,  
CC hypertensive, and vulnary. The albumin fusion protein nucleic acid may  
CC be used in gene therapy to treat disorders. The albumin fusion protein is  
CC useful for diagnosing, treating, preventing or ameliorating diseases or  
CC disorders comprising indication: Y. The diseases or disorders include:  
CC cancer (e.g. leukemia, colon, bone, breast, liver or lung cancer),  
CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute  
CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,  
CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme  
CC disease), reproductive system disorders (e.g. prostaticitis, inguinal  
CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-  
CC Leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,  
CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy  
CC or cachexia), cardiovascular disease (e.g. rhabdomyoma, heart disease,  
CC arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or  
CC hyponatremia), mixed foetal diseases (e.g. foetal alcohol syndrome,  
CC Down's syndrome, Patou syndrome, Turner's syndrome, Apert syndrome or Tay-  
CC -Sachs disease), excretory diseases (e.g. urinary incontinence, urinary  
CC tract infections or renal disorders), neural or sensory disease (e.g.  
CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,  
CC cerebellar ataxia, attention deficit disorder, autism or obsessive  
CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or  
CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's  
CC disease or glomerulonephritis), digestive diseases (e.g. portal  
CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)  
CC or connective tissue or epithelial diseases (e.g. Crohn's disease,  
CC scleroderma, wound healing or epidermolysis bullosa). This sequence  
CC represents a therapeutic protein X relating to the albumin fusion protein  
CC of the invention. The sequence listing data for this specification was  
CC downloaded from the USPTO website.  
XX  
XX  
SQ Sequence 403 AA;  
Query Match 100.0%; Score 1526; DB 8; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.8e-155;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QGGRREDGPGACYGFDLYFLIDKSGSVLHHMNEIYFVVEQLAHKFI SPQLRMSFIVFST 60  
DB 28 QGGRREDGPGACYGFDLYFLIDKSGSVLHHMNEIYFVVEQLAHKFI SPQLRMSFIVFST 87  
QY 61 RGTTLMLKLTEDRQIRQGLELQKVLPGGDTYVHSGFERASBOIYYENRGYRTASVITA 120  
DB 88 RGTTLMLKLTEDRQIRQGLELQKVLPGGDTYVHSGFERASBOIYYENRGYRTASVITA 147  
QY 121 LTGDELHEDLFFYSERANRSDLGAIYVGVKDFNEFETLARLADS KDHYFPVNDGFOA 180  
DB 148 LTGDELHEDLFFYSERANRSDLGAIYVGVKDFNEFETLARLADS KDHYFPVNDGFOA 207  
QY 161 LOGIHSILKKSCEIILAEPSITCAGESFOVYVGRNGFRHANNDRVLCSEFKINDSVTL 240

DB 208 LOGIHSILKKSCEIILAEPSITCAGESFOVYVGRNGFRHANNDRVLCSEFKINDSVTL 267  
QY 241 NEKPFSEVEDTYLLCPAPILKEVGKKAALQVSNNDGSEFISSVITTTTHCSGD 293  
DB 268 NEKPFSEVEDTYLLCPAPILKEVGKKAALQVSNNDGSEFISSVITTTTHCSGD 320  
RESULT 5  
AD100550  
ID AD100550 standard; protein; 551 AA.  
XX  
AC AD100550:  
XX  
DT 22-APR-2004 (first entry)  
XX  
XX Human TANGO 197 Ig fusion mutated protein - plaemid p0610.  
XX  
KM fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KM cutaneous; inhalation anthrax; human; TANGO 197 Ig fusion; mutant;  
KM Plaemid p0610; mutein.  
XX  
OS Homo sapiens.  
XX Synthetic.  
XX US2003144193-A1.  
XX  
PD 31-JUL-2003.  
XX  
PF 24-JUL-2002; 2002US-00201292.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-720708/68.  
DR N-PSDB; AD100549.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
XX to the vWF.  
XX  
PS Claim 45; SEQ ID NO 18; 86pp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
XX Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
XX acid sequence heterologous to the vWF. The polypeptide of the invention  
XX demonstrates antibacterial activities whilst the composition and method  
XX may be useful in preventing or ameliorating the symptoms of cutaneous  
XX and/or inhalation anthrax. The current sequence is that of the human  
XX TANGO 197 Ig mutated fusion protein of the invention.  
XX  
SQ Sequence 551 AA;  
Query Match 100.0%; Score 1526; DB 7; Length 551;  
Best Local Similarity 100.0%; Pred. No. 2.8e-155;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QGGRREDGPGACYGFDLYFLIDKSGSVLHHMNEIYFVVEQLAHKFI SPQLRMSFIVFST 60  
DB 28 QGGRREDGPGACYGFDLYFLIDKSGSVLHHMNEIYFVVEQLAHKFI SPQLRMSFIVFST 87  
QY 61 RGTTLMLKLTEDRQIRQGLELQKVLPGGDTYVHSGFERASBOIYYENRGYRTASVITA 120  
DB 88 RGTTLMLKLTEDRQIRQGLELQKVLPGGDTYVHSGFERASBOIYYENRGYRTASVITA 147  
QY 121 LTGDELHEDLFFYSERANRSDLGAIYVGVKDFNEFETLARLADS KDHYFPVNDGFOA 180

Db 148 LTDELHEDLFYFYSERANRSDIGAIVYCVGVDFNETQLARIADSKDHVPVNDGFQA 207  
 Qy 181 LQGIHSHILKKSCEIILAEPSTTCAGSPQVVRGNGFRHARVNDVLCSEFKINDSVTL 240  
 Db 208 LQGIHSHILKKSCEIILAEPSTTCAGSPQVVRGNGFRHARVNDVLCSEFKINDSVTL 267  
 Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 293  
 Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 320

## RESULT 6

ADM64576  
 ID ADM64576 standard; protein; 551 AA.

AC ADM64576;

DT 03-JUN-2004 (first entry)

DE Mouse TANGO197-immunoglobulin (Ig) fusion protein.

XX antibacterial; gene therapy;

KM von Willebrand factor A-like domain amino acid sequence;

KM vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;

KM inhalation anthrax; mouse; TANGO197; immunoglobulin; Ig; fusion protein;

gene.

XX Mus sp.

OS Synthetic.

XX US2003134786-A1.

XX 17-JUL-2003.

XX 20-DEC-2001; 2001US-00038307.

XX 20-DEC-2001; 2001US-00038307.

XX 20-DEC-2001; 2001US-00038307.

XX (ROTT/) ROTTMAN J B.

PA (OKEE/) O'KEEFE T L.

PA (OZKA/) OZKAYNAK E.

PA (HEAL/) HEALEY J J.

XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

XX WPI; 2003-829643/77.

XX N-PSDB; ADM64575.

XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or

PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like

PT domain (vWF) amino acid sequence and an amino acid sequence heterologous

PT to the vWF.

XX Claim 44; SEQ ID NO 18; 649p; English.

XX The invention describes a fusion polypeptide comprising a von Willebrand

CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence

CC heterologous to the vWF. Also described are: a method of preventing or

CC ameliorating a symptom of anthrax in a subject thought to be at risk for

CC exposure to or suspected of having been exposed to Bacillus anthracis;

CC and a pharmaceutical composition comprising the novel fusion polypeptide.

CC The composition and method are useful in preventing or ameliorating

CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid

CC sequence of a fusion protein comprising mouse TANGO197 and immunoglobulin

CC (Ig) that can be used to treat exposure to or prevent a symptom of

CC anthrax.

XX Sequence 551 AA;

XX Query Match 100.0%; Score 1526; DB 7; Length 551;

XX Best Local Similarity 100.0%; Pred. NO. 2.8e-155;

XX Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGRRREGGPA CYGGFDLYFLDKSGSVLHHWNEIYYFVEQLAHKTSPLQKMSFIYFST 60  
 Db 28 QGRRREGGPA CYGGFDLYFLDKSGSVLHHWNEIYYFVEQLAHKTSPLQKMSFIYFST 87  
 Qy 61 RGTLMKLTERRBOIRGLELQVLPQSDTYMEGGERASEQIYENRQGYRTASVITA 120  
 Db 88 RGTLMKLTERRBOIRGLELQVLPQSDTYMEGGERASEQIYENRQGYRTASVITA 147  
 Qy 121 LTDELHEDLFYFYSERANRSDIGAIVYCVGVDFNETQLARIADSKDHVPVNDGFQA 180  
 Db 148 LTDELHEDLFYFYSERANRSDIGAIVYCVGVDFNETQLARIADSKDHVPVNDGFQA 207  
 Qy 181 LQGIHSHILKKSCEIILAEPSTTCAGSPQVVRGNGFRHARVNDVLCSEFKINDSVTL 240  
 Db 208 LQGIHSHILKKSCEIILAEPSTTCAGSPQVVRGNGFRHARVNDVLCSEFKINDSVTL 267  
 Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 293  
 Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 320

## RESULT 7

ABB90750  
 ID ABB90750 standard; protein; 564 AA.

AC ABB90750;

DT 30-MAY-2002 (first entry)

DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 232.

XX Human; mouse; rat; TEM; tumour endothelial marker; TEM; PEM; cytostatic;

KM normal endothelial marker; pan-endothelial marker; immunostimulant;

KM antiangiogenic; tumour; neovascularisation; vascularised tumour;

KM polygenic kidney disease; diabetes; retinopathy; rheumatoid arthritis;

KM psoriasis.

XX Homo sapiens.

XX WO200210217-A2.

XX 07-FEB-2002.

XX 01-AUG-2001; 2001WO-US024031.

XX 02-AUG-2000; 2000US-0222599P.

XX 11-AUG-2000; 2000US-0224360P.

XX 11-APR-2001; 2001US-0282850P.

XX (UYJO ) UNIV JOHNS HOPKINS.

XX St Croix B, Kinzler KW, Vogelstein B;

XX WPI; 2002-291856/33.

XX N-PSDB; ABL92104.

XX An isolated molecule comprising an antibody variable region which

PT specifically binds to an extracellular domain of a tumor endothelial

PT marker (TEM) protein, useful for inhibiting tumor growth.

XX Claim 1; Page 209-210; 331p; English.

XX The invention relates to an isolated molecule comprising an antibody

CC variable region which specifically binds to an extracellular domain of a

CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,

CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM

CC proteins have cytostatic, immunostimulant and antiangiogenic activity.

CC They are useful for inhibiting tumour growth, neovascularisation in subjects

CC bearing a vascularised tumour, polycystic kidney disease, diabetic

CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM

CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)

CC are disclosed, as are marker oligonucleotide sequences: tumour

CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal

CC

CC

CC	endothelial markers (NEM)	ABL92042-ABL92074; and pan-endothelial markers (PEM)	ABL91903-ABL91995
XX	Sequence	564 AA;	
XX	Query Match	100.0%;	Pred. 1526; DB 5; Length 564;
XX	Best Local Similarity	100.0%;	Pred. No. 2.9e-155;
XX	Matches	293;	Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1	QGGRRBDGPGACVCGFDLYFIIDKSGSVLHNNELIYFVEQLAHKFTISPOLRMSPIVST	60
DB	28	QGGRRBDGPGACVCGFDLYFIIDKSGSVLHNNELIYFVEQLAHKFTISPOLRMSPIVST	87
QY	61	RGTTLMKLTEDREQIRQGLLELOKVLPGSDTTMHSGFERASQIYYENRQGRFASVITA	120
DB	88	RGTTLMKLTEDREQIRQGLLELOKVLPGSDTTMHSGFERASQIYYENRQGRFASVITA	147
QY	121	LITGEIHEDLFFPSSEBANSRLGAIYCVGKDPNENQLARINDSKDHPFVNDGFOA	180
DB	148	LITGEIHEDLFFPSSEBANSRLGAIYCVGKDPNENQLARINDSKDHPFVNDGFOA	207
QY	181	LOGIITHSLKKSCETILAEPTSTCAGESFOVVRGNGCFRHRANDRYLCSFKINDSVTL	240
DB	208	LOGIITHSLKKSCETILAEPTSTCAGESFOVVRGNGCFRHRANDRYLCSFKINDSVTL	267
QY	241	NEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSPFISSVITTTTCSDG	293
DB	268	NEKPSVEDTYLLCPAPILKEVGKKAALQVSNMDGLSPFISSVITTTTCSDG	320
RESULT 8			
ID	ABB90724	standard; protein; 564 AA.	
AC	ABB90724;		
DT	30-MAY-2002	(first entry)	
XX	Human Tumour Endothelial Marker polypeptide SEQ ID NO 187.		
XX	Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic;		
XX	normal endothelial marker; pan-endothelial marker; immunostimulant;		
XX	antiangiogenic; tumour; neovascularogenesis; vascularised tumour;		
XX	polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;		
XX	psoriasis.		
OS	Homo sapiens.		
PN	MO200210217-A2.		
PD	07-FEB-2002.		
PE	01-AUG-2001; 2001MO-US024031.		
PR	02-AUG-2000; 2000US-0222599P.		
PR	11-AUG-2000; 2000US-0224360P.		
PR	11-APR-2001; 2001US-0282850P.		
PA	(UYJO ) UNIV JOHNS HOPKINS.		
PI	St Croix B, Kinzler KW, Vogelstein B;		
DR	WPI; 2002-291856/33.		
XX	An isolated molecule comprising an antibody variable region which		
PT	specifically binds to an extracellular domain of a tumor endothelial		
PT	marker (TEM) protein, useful for inhibiting tumor growth.		
PS	Disclosure; Page 136-137; 33pp; English.		
CC	The invention relates to an isolated molecule comprising an antibody		
CC	variable region which specifically binds to an extracellular domain of a		
CC	tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,		

CC	ABB90749	ABB90750 and ABB90769.	The antibodies which bind to TEM proteins have cytostatic, immunostimulant and antitumorigenic activity.
CC	They are useful	for inhibiting tumour growth, neovascularogenesis in subjects bearing a vascularised tumour, polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis and psoriasis.	Human, mouse and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789) are disclosed, as are marker oligonucleotide sequences; tumour endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers (PEM) ABL91903-ABL91995
CC	CC	CC	CC
XX	SQ	Sequence 564 AA;	
Query Match	100.0%;	Score 1526;	DB 5; Length 564;
Best Local Similarity	100.0%;	Pred. No. 2.9e-155;	
Matches 293;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0
OY	1	QGRRREDGPACYGCFDLFYFLDKSGSVLHHMNEIYYVEQLAHKFIISPOLRMSFIVFST	60
Db	28	QGRRREDGPGCYGGFDLYFLDSSGSVLHMMNEIYYVEQLAHKFISPOLRMSFIVST	87
OY	61	RGTLLMLKTEDREQIRGLLELOKVLPQGDVTYMHGEFERASEQIYENRGYRTASVIA	120
Db	88	RGTLLMLKTEDREQIRGLEELQKVLPGSDTYMHGFERASEQIYENRGYRTASVIA	147
OY	121	LTDGELHEDLFYYSEREANRSRDYGAIYYCVGKD FNETQLARINDSKOHFFPVNDGFOA	180
Db	148	LTDGELHEDLFYYSEREANRSRDYGAIYYCVGVDFNETQLARINDSKOHFFPVNDGFOA	207
OY	181	LOGIIHSILTKKSCIEILAABESTICAGSFPVVYRGNGFRHARNDRVLCFSKINDSVTL	240
Db	208	LOGIIHSILTKKSCIEILAABESTICAGSFVVYRGNGFRHARNDRVLCFSKINDSVTL	267
OY	241	NEKPFSVEDTTLCPAPILKEVGKKAALQVMNDGLSISSVITTTTHCSDG	293
Db	268	NEKPFSVEDTTLCPAPILKEVGKKAALQVMNDGLSISSVITTTTHCSDG	320
RESULT 9			
ABP54904	ID	ABP54904 standard; protein; 564 AA.	
XX	ABP54904;		
XX	08-JAN-2003	(first entry)	
DT	XX	Human anthrax toxin receptor.	
DE	XX	Anthrax; toxin; receptor; human; TEM6; antibacterial.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	Peptide	1..27	/label= Signal_peptide
FT	Protein	28..564	/label= Mature_protein
FT	Domain	28..320	/note= "extracellular domain"
FT	Domain	44..216	/note= "von Willebrand factor A domain"
FT	Domain	320..343	/note= "putative transmembrane domain"
FT	Domain	344..564	/note= "cytoplasmic domain"
PN	WO200246228-A2.		
PD	13-JUN-2002.		
PF	03-OCT-2001; 2001WO-US030941.		
PR	05-DEC-2000; 2000US-0251481P.		



XX (WISC) WISCONSIN ALUMNI RES FOUND.  
 PA Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
 PI WPI; 2002-713235/77.  
 XX N-PSDB; ABV73882.  
 DR Novel isolated polypeptide useful for identifying agent that prevents or  
 PT reduces effect of anthrax toxin on host cell, for treating human or non-  
 PT human animal suffering from anthrax.  
 XX  
 PS Claim 1; Page 37-39; 45pp; English.  
 XX  
 CC The present sequence is the protein sequence of a human surface-bound  
 CC anthrax toxin receptor (ATR), previously designated TEM8, and identified  
 CC following a database screening using a newly isolated human surface-bound  
 CC ATR (see ABP54903). The present sequence differs from the newly isolated  
 CC human ATR only in the cytoplasmic domain (221 rather than 25 amino acids  
 CC long), suggesting differential splicing of a primary mRNA transcript.  
 CC Anthrax toxin protective antigen (PA) binds to these ATRs at a von  
 CC Willebrand factor A domain located in the extracellular domain of ATR.  
 CC The invention provides ATR polypeptides and polynucleotides, vectors,  
 CC host cells, and transgenic and knock-out animals. It also provides  
 CC methods for identifying molecules that bind the ATR and which reduce the  
 CC toxicity of anthrax toxin. A claimed method for treating anthrax in a  
 CC human or animal involves administering an agent that inhibits binding  
 CC between PA and ATR at a level effective to reduce the severity of  
 CC anthrax. Suitable agents include the present polypeptide or a PA-binding  
 CC fragment of it, a PA-binding polypeptide at least 80% identical to these,  
 CC a fusion protein, a monoclonal or polyclonal antibody, a polysaccharide,  
 CC a lipid or a nucleic acid  
 XX  
 SQ Sequence 564 AA;  
 Query Match 100.0%; Score 1526; DB 5; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-155;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QGGRREDGPGACYGFDLYFLDKSGSVLHHMNEIYFVEQLAKHFTSPQLRMSFIYFST 60  
 DB 28 QGGRREDGPGACYGFDLYFLDKSGSVLHHMNEIYFVEQLAKHFTSPQLRMSFIYFST 87  
 QY 61 RGTTLMLKLTEDREIROGLLELOKVLPGSDTYMHGFERASEQIYENRQGYRTASVYIA 120  
 DB 88 RGTTLMLKLTEDREIROGLLELOKVLPGSDTYMHGFERASEQIYENRQGYRTASVYIA 147  
 QY 121 LTDELHEDLFFYSERANRSRDIGAIYVCVGDVFNETQLARIADSKDHPVNDGFOA 180  
 DB 148 LTDELHEDLFFYSERANRSRDIGAIYVCVGDVFNETQLARIADSKDHPVNDGFOA 207  
 QY 181 LOGIHSILKSKCIEILAAEPSTICAGESFOVVVRNGGFHARVNDVLTCSFKINDSVTL 240  
 DB 208 LOGIHSILKSKCIEILAAEPSTICAGESFOVVVRNGGFHARVNDVLTCSFKINDSVTL 267  
 QY 241 NEKPFSEVDYLLCPAPILKEVGKAAALQVSMNDGLSFISSVYITTTTHCSGD 293  
 DB 268 NEKPFSEVDYLLCPAPILKEVGKAAALQVSMNDGLSFISSVYITTTTHCSGD 320

RESULT 10  
 ABUS4457  
 ID ABUS4457 standard; protein; 564 AA.  
 XX  
 AC ABUS4457;  
 XX  
 DT 12-MAR-2003 (first entry)  
 XX  
 DE Human tumour endothelial marker TEM 19.  
 XX  
 KW Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;  
 KW tumour endothelial marker; normal endothelial marker; PEM;  
 KW pan-endothelial marker; polycystic kidney disease; psoriasis;

KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;  
 KW neovascularization; immune response; cytotoxic; antidiabetic;  
 KW opthalmological; antineumatic; antiarthritic; antipsoriatic.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200283874-A2.  
 XX  
 PD 24-OCT-2002.  
 XX  
 XX 10-APR-2002; 2002WO-US008253.  
 PE  
 PR 11-APR-2001; 2001US-0282850P.  
 PR 06-FEB-2002; 2002US-0354262P.  
 XX  
 PA (UJVO) UNIV JOHNS HOPKINS.  
 PI Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;  
 DR WPI; 2003-093016/08.  
 XX N-PSDB; ABX72029.  
 XX  
 PT New purified human transmembrane protein, designated as tumor endothelial  
 PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors,  
 PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or  
 PT psoriasis.  
 XX  
 PS Disclosure; Page 226-227; 374pp; English.  
 XX  
 CC The present invention relates to a novel method for the isolation of  
 CC endothelial cells (ECs), and the identification of genes expressed in  
 CC normal and tumour ECs. Tumour endothelial marker (TEM), normal  
 CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are  
 CC identified in human ECs. The human EC marker proteins and the  
 CC polynucleotide sequences encoding them are useful for detecting,  
 CC diagnosing or treating tumours as well as polycystic kidney disease,  
 CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also  
 CC useful for inhibiting neovascularization or tumour angiogenesis, for  
 CC inducing an immune response to tumour endothelial cells in a patient, or  
 CC for identifying candidate drugs for treating tumours. The present  
 CC sequence represents a human TEM or NEM protein of the invention  
 XX  
 SQ Sequence 564 AA;  
 Query Match 100.0%; Score 1526; DB 6; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-155;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QGGRREDGPGACYGFDLYFLDKSGSVLHHMNEIYFVEQLAKHFTSPQLRMSFIYFST 60  
 DB 28 QGGRREDGPGACYGFDLYFLDKSGSVLHHMNEIYFVEQLAKHFTSPQLRMSFIYFST 87  
 QY 61 RGTTLMLKLTEDREIROGLLELOKVLPGSDTYMHGFERASEQIYENRQGYRTASVYIA 120  
 DB 88 RGTTLMLKLTEDREIROGLLELOKVLPGSDTYMHGFERASEQIYENRQGYRTASVYIA 147  
 QY 121 LTDELHEDLFFYSERANRSRDIGAIYVCVGDVFNETQLARIADSKDHPVNDGFOA 180  
 DB 148 LTDELHEDLFFYSERANRSRDIGAIYVCVGDVFNETQLARIADSKDHPVNDGFOA 207  
 QY 181 LOGIHSILKSKCIEILAAEPSTICAGESFOVVVRNGGFHARVNDVLTCSFKINDSVTL 240  
 DB 208 LOGIHSILKSKCIEILAAEPSTICAGESFOVVVRNGGFHARVNDVLTCSFKINDSVTL 267  
 QY 241 NEKPFSEVDYLLCPAPILKEVGKAAALQVSMNDGLSFISSVYITTTTHCSGD 293  
 DB 268 NEKPFSEVDYLLCPAPILKEVGKAAALQVSMNDGLSFISSVYITTTTHCSGD 320

RESULT 11  
 ABUS4431  
 ID ABUS4431 standard; protein; 564 AA.  
 XX

AC ABUS4431;  
XX  
XX 12-MAR-2003 (first entry)  
XX  
XX Human tumour endothelial marker TEM 8.  
DE  
XX Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;  
KM Tumour endothelial marker; normal endothelial marker; PEM;  
KM pan-endothelial marker; polycystic kidney disease; psoriasis;  
KM diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;  
KM neovascularization; immune response; cytostatic; antidiabetic;  
KM ophthalmological; antineoplastic; antirheumatic; antipsoriatic.  
XX  
OS Homo sapiens.  
XX  
PN W0200283874-A2.  
XX  
XX 24-OCT-2002.  
XX  
XX 10-APR-2002; 2002WO-US008253.  
XX  
XX 11-APR-2001; 2001US-0282850P.  
PR 06-FEB-2002; 2002US-0354262P.  
XX  
XX (UYJO ) UNIV JOHNS HOPKINS.  
XX  
XX Carson-Walter E, St Croix B, Kinzler KM, Vogelstein B;  
PI  
XX WPI; 2003-093016/08.  
XX  
XX New purified human transmembrane protein, designated as tumor endothelial  
PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors,  
PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or  
PT psoriasis.  
XX  
XX Disclosure; Page 138-139; 374pp; English.  
XX  
XX The present invention relates to a novel method for the isolation of  
CC endothelial cells (ECs), and the identification of genes expressed in  
CC normal and tumor ECs. Tumour endothelial marker (TEM), normal  
CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are  
CC identified in human ECs. The human EC marker proteins and the  
CC polynucleotide sequences encoding them are useful for detecting,  
CC diagnosing or treating tumours as well as polycystic kidney disease,  
CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also  
CC useful for inhibiting neovascularization or tumour angiogenesis, for  
CC inducing an immune response to tumour endothelial cells in a patient, or  
CC for identifying candidate drugs for treating tumours. The present  
CC sequence represents a human TEM or NEM protein of the invention  
XX  
SQ Sequence 564 AA;  
Query Match 100.0%; Score 1526; DB 6; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.9e-155;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QGGRREDGPGACGFDLYFLIDKSGSVLHNNETIYFVQOLAKHFISSPOLRMSFIYEST 60  
DB 28 QGGRREDGPGACGFDLYFLIDKSGSVLHNNETIYFVQOLAKHFISSPOLRMSFIYEST 87  
QY 61 RGTILMKLTEDREQIRQGLBELQKVLPGDITVMEHGEFRASEQIYENRQGYRTASVITA 120  
DB 88 RGTILMKLTEDREQIRQGLBELQKVLPGDITVMEHGEFRASEQIYENRQGYRTASVITA 147  
QY 121 LTDEHLEDLFFYGEREARRSRDLGATVYCVGVDFNETOLARLADSDHVPVNDGQQA 180  
DB 148 LTDEHLEDLFFYGEREARRSRDLGATVYCVGVDFNETOLARLADSDHVPVNDGQQA 207  
QY 181 LQGIHSLIKSKSCEIILAAEPSTICAGSPVVVNGNFRARVNDVLCGFKINDSVTL 240  
DB 208 LQGIHSLIKSKSCEIILAAEPSTICAGSPVVVNGNFRARVNDVLCGFKINDSVTL 267  
QY 241 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSDG 293

DB 268 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSDG 320  
RESULT 12  
ADJ70017  
ID ADJ70017 standard; protein; 564 AA.  
XX  
XX ADJ70017;  
AC  
XX 06-MAY-2004 (first entry)  
DT  
XX Human heat mitochondrial protein as a therapeutic target SeqID1823.  
DE  
XX  
XX mitochondrial; human; screening assay; diabetes mellitus;  
KM Huntington's disease; osteoarthritis;  
KM Leber's hereditary optic neuropathy; LHON;  
KM mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
KM myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
KM neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;  
KM osteopathic; ophthalmological; cytostatic.  
XX  
OS Homo sapiens.  
XX  
XX W02003087768-A2.  
PN  
XX 23-OCT-2003.  
XX  
XX 04-APR-2003; 2003WO-US010870.  
XX  
XX 12-APR-2002; 2002US-0372843P.  
PR 17-JUN-2002; 2002US-038987P.  
XX 20-SEP-2002; 2002US-0412418P.  
XX  
XX (MITO-) MITOKOR.  
PA (BUCK-) BUCK INST AGE RES.  
PA  
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GW;  
PI Wainrock DE;  
XX  
XX WPI; 2003-845369/78.  
DR  
XX Identifying a mitochondrial target for drug screening assays and for  
XX treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.  
XX  
XX Claim 1; SEQ ID NO 1823; 180pp; English.  
PS  
XX  
XX This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, nootropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytoskeletal activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
SQ Sequence 564 AA;  
Query Match 100.0%; Score 1526; DB 7; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.9e-155;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QGGRREDGPGACGFDLYFLIDKSGSVLHNNETIYFVQOLAKHFISSPOLRMSFIYEST 60

Db 28 QGRRBDGPGACVGFPLDYFLDKSGSVLHMNNEYFVEQLAHKFTSPQLRMSFVST 87  
 Qy 61 RGTTLMLKLTEDREQIRQGLEBELQKVLPGDPTVMHEGFERASBOIYYENRQGYRTASV11A 120  
 Db 88 RGTTLMLKLTEDREQIRQGLEBELQKVLPGDPTVMHEGFERASBOIYYENRQGYRTASV11A 147  
 Qy 121 LTDELHEDLFFYSERANRSDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOA 180  
 Db 148 LTDELHEDLFFYSERANRSDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOA 207  
 Qy 181 LOGIHSILKKSCEIILAAEPSTICAGESFOVVVNGNFRHARVNDVLCSPKINDSVTL 240  
 Db 208 LOGIHSILKKSCEIILAAEPSTICAGESFOVVVNGNFRHARVNDVLCSPKINDSVTL 267  
 Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 293  
 Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 320  
 RESULT 13  
 ADR48216 ID ADR48216 standard; protein; 564 AA.  
 AC ADR48216,  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human tumour endothelial marker 8 precursor protein SEQ:4.  
 XX  
 KW pancreatic cancer-associated transcript; pancreatic cancer; human;  
 KW cytostatic; gene therapy; protein therapy;  
 KW tumour endothelial marker 8 precursor; TEM8.  
 XX  
 OS Homo sapiens.  
 PN WC02004074510-A1.  
 XX  
 PD 02-SEP-2004.  
 XX  
 PF 18-FEB-2004; 2004WO-AU000194.  
 XX  
 PR 18-FEB-2003; 2003AU-00900747.  
 XX  
 PA (GARV-) GARVAN INST MEDICAL RES.  
 PI Blankin A, Segara D, Henshall S, Sutherland R;  
 XX  
 DR WPI, 2004-635591/61.  
 DR N-PSDB; ADR48215.  
 XX  
 PT Detecting pancreatic cancer-associated transcript in a biological sample,  
 PT useful for diagnosing or treating the disease, comprises contacting the  
 PT sample with a polynucleotide that selectively hybridizes to a specific  
 FT sequence.  
 XX  
 PS Claim 70; SEQ ID NO 4; 263bp; English.  
 XX  
 CC The present invention describes a method for detecting a pancreatic  
 CC cancer-associated transcript in a biological sample. The method comprises  
 CC contacting the biological sample with a polynucleotide that selectively  
 CC hybridizes to a sequence at least 80% identical to a sequence as shown in  
 CC any one of Tables 3 to 25 in the specification or having the GenBank  
 CC Accession Number AF279145. Also described: (1) diagnosing pancreatic  
 CC cancer in a human or animal subject being tested, determining the  
 CC likelihood that a subject having a pancreatic cancer will survive, or  
 CC determining the suitability of a subject having a pancreatic cancer for  
 CC surgical resection therapy; (2) detecting a pancreatic cancer-associated  
 CC polypeptide in a biological sample; (3) determining the likelihood that a  
 CC subject having a pancreatic cancer will survive; and (4) monitoring the  
 CC efficacy of a therapeutic treatment of pancreatic cancer. A pancreatic  
 CC cancer-associated transcript has cytostatic activity, and can be used in  
 CC gene and protein therapy. A pancreatic cancer-associated transcript  
 CC polynucleotide, a vector comprising the polynucleotide, an isolated

CC polypeptide or an antibody that binds to the isolated polypeptide can be  
 CC used for diagnosing or prognosing pancreatic cancer or for preparing a  
 CC medicament for the treatment of pancreatic cancer. The prognostic or  
 CC diagnostic methods are useful for the early detection of pancreatic  
 CC cancer or its metastases, and for monitoring the progress of disease such  
 CC as during remission or following surgery or chemotherapy. The present  
 CC sequence represents human tumour endothelial marker 8 precursor (TEM8),  
 CC which is used in the exemplification of the present invention.  
 XX  
 SO Sequence 564 AA;  
 Query Match 100.0%; Score 1526; DB 8; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-155;  
 Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 QGRRBDGPGACVGFPLDYFLDKSGSVLHMNNEYFVEQLAHKFTSPQLRMSFVST 60  
 Db 28 QGRRBDGPGACVGFPLDYFLDKSGSVLHMNNEYFVEQLAHKFTSPQLRMSFVST 87  
 Qy 61 RGTTLMLKLTEDREQIRQGLEBELQKVLPGDPTVMHEGFERASBOIYYENRQGYRTASV11A 120  
 Db 88 RGTTLMLKLTEDREQIRQGLEBELQKVLPGDPTVMHEGFERASBOIYYENRQGYRTASV11A 147  
 Qy 121 LTDELHEDLFFYSERANRSDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOA 180  
 Db 148 LTDELHEDLFFYSERANRSDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOA 207  
 Qy 181 LOGIHSILKKSCEIILAAEPSTICAGESFOVVVNGNFRHARVNDVLCSPKINDSVTL 240  
 Db 208 LOGIHSILKKSCEIILAAEPSTICAGESFOVVVNGNFRHARVNDVLCSPKINDSVTL 267  
 Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 293  
 Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSVITTTTHCSGD 320  
 RESULT 14  
 AAB01469 ID AAB01469 standard; protein; 403 AA.  
 XX  
 AC AAB01469;  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE Human gene 4 encoded secreted protein HMLFR02, SEQ ID NO:125.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification;  
 KW chromosome 19.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Protein /label= signal\_peptide  
 FT Misc-difference 28..403  
 FT /note= "Mature human secreted protein"  
 FT Misc-difference 175  
 FT /label= Unknown  
 FT /note= "Encoded by GKT"  
 FT Misc-difference 320  
 FT /label= Unknown  
 FT /note= "Encoded by SGT"  
 FT Misc-difference 331  
 FT /label= Unknown

FT /note= "Encoded by KTC"  
FT Mlec-difference 368  
FT /label= Unknown  
FT /note= "Encoded by WGC"  
XX  
XX WO200134626-A1.  
XX  
XX 17-MAY-2001.  
XX  
XX 01-NOV-2000; 2000WO-US030045.  
XX  
XX 05-NOV-1999; 99US-0163581P.  
XX 30-JUN-2000; 2000US-0215133P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsu JI GA, Moore PA, Birse CE, Ni J;  
XX WPI; 2001-308778/32.  
XX N-PSDB; AAD05334.  
XX  
XX New nucleic acid molecules encoding 28 human secreted proteins for  
XX diagnosing, preventing, treating or ameliorating medical conditions and  
XX used as food additives or preservatives.  
XX  
XX Claim 11; Page 505-506; 562pp; English.  
XX  
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
XX protein genes, and AAE01435-AAE01513 represent the proteins they encode.  
XX AAE01514-AAE01544 represent human secreted protein fragments or variants.  
XX The genes and their secreted proteins are useful for preventing, treating  
XX or ameliorating medical conditions, e.g., by protein or gene therapy.  
XX Pathological conditions can be diagnosed by determining the amount of the  
XX new protein in a sample or by determining the presence of mutations in  
XX the new genes. Specific uses are described for each of the 28 genes,  
XX based on the tissues in which they are most highly expressed, and include  
XX developing products for the diagnosis or treatment of proliferative  
XX disorders, cancer, tumours, foetal and developmental abnormalities,  
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
XX angiotensin disorders, kidney disorders, gastrointestinal disorders,  
XX pregnancy-related disorders, endocrine disorders, and infections. The  
XX proteins can also be used to aid wound healing and epithelial cell  
XX proliferation, to prevent skin aging due to sunburn, to maintain organs  
XX before transplantation, for supporting cell culture of primary tissues,  
XX to regenerate tissues, to identify their cognate ligands or binding  
XX partners, and in chemotaxis, and can be used as a food additive or  
XX preservative to modify storage properties. Antibodies specific for a  
XX protein of the invention can be used in alleviating symptoms associated  
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
XX present sequence represents a human secreted protein of the invention  
XX  
XX Sequence 403 AA;  
SQ  
Query Match 99.3%; Score 1515; DB 4; Length 403;  
Best Local Similarity 99.7%; Pred. No. 2,7e-154; Indels 0; Gaps 0;  
Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 OGGRRDEGPGACGAGFDLYFINDKSGSVLHMHNEIYYVEVQLAHKFIQPMRSTIVST 60  
DB 28 OGGRRDEGPGACGAGFDLYFINDKSGSVLHMHNEIYYVEVQLAHKFIQPMRSTIVST 87  
QY 61 RGTTLMLKLTEDREQIRQGLEBELQKVLPGGDIYMHGEPFASBQIYYENRQGRITASVITA 120  
DB 88 RGTTLMLKLTEDREQIRQGLEBELQKVLPGGDIYMHGEPFASBQIYYENRQGRITASVITA 147  
QY 121 LTDDGHLHDLFPYSSREANRSPDLGAIYCVGVKDPNPTOLARLADSDHYFPVNDGQQA 180  
DB 148 LTDDGHLHDLFPYSSREANRSPDLGAIYCVGVKDPNPTOLARLADSDHYFPVNDGQQA 207

QY 181 LOGIHSILKSKSCIEILAAEPSTICAGESFOVVRGNGFRRHARNDVRLCSFKINDSVTL 240  
DB 208 LOGIHSILKSKSCIEILAAEPSTICAGESFOVVRGNGFRRHARNDVRLCSFKINDSVTL 267  
QY 241 NEKRFVSVEDTYLLCPAPILKEVGKMAALQVSNMDGSLFISSSVITTTTHCSD 292  
DB 268 NEKRFVSVEDTYLLCPAPILKEVGKMAALQVSNMDGSLFISSSVITTTTHCSD 319  
RESULT 15  
ABG63873 standard; protein; 403 AA.  
XX  
XX ABG63873;  
AC  
XX 27-AUG-2002 (first entry)  
DT  
XX  
XX Human albumin fusion protein #548.  
DE  
XX Albumin fusion protein; therapeutic protein X; human albumin; HA;  
XX human serum albumin; HSA; cancer; reproductive disorder;  
XX digestive disorder; immune disorder; endocrine disorder;  
XX haematopoietic disorder; neural disorder; connective disorder;  
XX cytostatic; antifertility; antiinflammatory; anticancer;  
XX immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;  
XX neuroprotective; antiParkinsonian; antimicrobial; neuroleptic;  
XX osteopathic; antiarthritic.  
XX  
XX Homo sapiens.  
OS  
XX Synthetic.  
OS  
XX WO200177137-A1.  
XX  
XX 18-OCT-2001.  
PD  
XX  
XX 12-APR-2001; 2001WO-US011988.  
PF  
XX  
XX 12-APR-2000; 2000US-0229358P.  
PR 25-APR-2000; 2000US-0199384P.  
PR 21-DEC-2000; 2000US-0256931P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Haseltine WA;  
XX  
XX WPI; 2002-010866/01.  
XX  
XX New fusion protein for treating disease e.g. diabetes comprises an  
XX albumin fused to a therapeutic protein.  
XX  
XX Claim 1; Page 872-873; 2102pp; English.  
XX  
XX The present invention relates to albumin fusion proteins comprising a  
XX therapeutic protein X and human albumin (HA, also known as human serum  
XX albumin, HSA). The proteins are useful for treating a disease or disorder  
XX that may be modulated by therapeutic protein X. The albumin extends the  
XX shelf-life of protein X, and may increase its biological activity in vivo  
XX activity. The protein is useful for treating and diagnosing disorders  
XX such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
XX disease, ulcerative colitis), immune disorders (e.g. acquired  
XX immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
XX haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
XX Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
XX schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
XX  
XX ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX  
XX Sequence 403 AA;  
SQ  
Query Match 99.3%; Score 1515; DB 5; Length 403;  
Best Local Similarity 99.7%; Pred. No. 2,7e-154; Indels 0; Gaps 0;  
Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 QGRRDGGPACYGFDLYFTLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVFST 60
Db 28 QGRRDGGPACYGFDLYFTLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVFST 87
Qy 61 RGTTLMLKLTEDREQIROGLAEIQKVLPGGDTYMHGFERASEQIYENRQGYRTASVITA 120
Db 88 RGTTLMLKLTEDREQIROGLAEIQKVLPGGDTYMHGFERASEQIYENRQGYRTASVITA 147
Qy 121 LTDEGLHEDLFPYSERREANRSRDIGAIYVCVGDQFNETOLARIADSKOHVPVNDGFOA 180
Db 148 LTDEGLHEDLFPYSERREANRSRDIGAIYVCVGDQFNETOLARIADSKOHVPVNDGFOA 207
Qy 181 LOGIHSILKSCIEIILAEPSSTICAGSFQVYVVRGNGFRHARVNDRVLCSPKINDSVTL 240
Db 208 LOGIHSILKSCIEIILAEPSSTICAGSFQVYVVRGNGFRHARVNDRVLCSPKINDSVTL 267
Qy 241 NEKPFVVEDTYLCPAPILKEVGMAALQVSMNDGLSFISSSVITTTTHCSD 292
Db 268 NEKPFVVEDTYLCPAPILKEVGMAALQVSMNDGLSFISSSVITTTTHCSD 319

```

Search completed: June 13, 2005, 19:56:50  
 Job time : 112.48 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using SW model

Run on: June 13, 2005, 19:46:02 ; Search time 28.2879 Seconds  
(without alignments)  
773.200 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_28\_320

Perfect score: 1526

Sequence: 1 QGRRDEGPGACYGFDLYF.....DGLSFSSSVITTTGCSGDG 293

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfltest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166	10.9	1155	1 US-08-286-889-46	Sequence 46, Appl
2	166	10.9	1155	1 US-08-485-618-46	Sequence 46, Appl
3	166	10.9	1155	1 US-08-362-652-46	Sequence 46, Appl
4	166	10.9	1155	2 US-08-605-672-46	Sequence 46, Appl
5	166	10.9	1155	2 US-08-482-293A-46	Sequence 46, Appl
6	166	10.9	1155	2 US-08-943-363-46	Sequence 46, Appl
7	166	10.9	1155	3 US-09-193-043-46	Sequence 46, Appl
8	166	10.9	1155	4 US-09-688-307A-46	Sequence 46, Appl
9	166	10.9	1155	4 US-09-350-259-46	Sequence 46, Appl
10	166	10.9	1161	1 US-08-485-618-53	Sequence 53, Appl
11	166	10.9	1161	1 US-08-362-652-53	Sequence 53, Appl
12	166	10.9	1161	2 US-08-605-672-53	Sequence 53, Appl
13	166	10.9	1161	2 US-08-482-293A-53	Sequence 53, Appl
14	166	10.9	1161	3 US-08-943-363-53	Sequence 53, Appl
15	166	10.9	1161	3 US-09-193-043-53	Sequence 53, Appl
16	166	10.9	1161	4 US-09-688-307A-53	Sequence 53, Appl
17	166	10.9	1161	4 US-09-350-259-53	Sequence 53, Appl
18	159.5	10.5	1151	1 US-08-485-618-37	Sequence 37, Appl
19	159.5	10.5	1151	1 US-08-362-652-37	Sequence 37, Appl
20	159.5	10.5	1151	1 US-08-605-672-37	Sequence 37, Appl
21	159.5	10.5	1151	2 US-08-482-293A-37	Sequence 37, Appl
22	159.5	10.5	1151	2 US-08-943-363-37	Sequence 37, Appl
23	159.5	10.5	1151	3 US-09-193-043-37	Sequence 37, Appl
24	159.5	10.5	1151	4 US-09-688-307A-37	Sequence 37, Appl
25	159.5	10.5	1151	4 US-09-350-259-37	Sequence 37, Appl
26	159.5	10.5	1151	4 US-08-485-618-55	Sequence 55, Appl
27	159.5	10.5	1161	1 US-08-485-618-55	Sequence 55, Appl

28	159.5	10.5	1161	1 US-08-362-652-55	Sequence 55, Appl
29	159.5	10.5	1161	2 US-08-605-672-55	Sequence 55, Appl
30	159.5	10.5	1161	2 US-08-482-293A-55	Sequence 55, Appl
31	159.5	10.5	1161	2 US-08-943-363-55	Sequence 55, Appl
32	159.5	10.5	1161	3 US-09-193-043-55	Sequence 55, Appl
33	159.5	10.5	1161	4 US-09-688-307A-55	Sequence 55, Appl
34	159.5	10.5	1161	4 US-09-350-259-55	Sequence 55, Appl
35	156	10.2	1161	1 US-08-173-457-2	Sequence 2, Appl
36	156	10.2	1161	1 US-08-286-889-2	Sequence 2, Appl
37	156	10.2	1161	1 US-08-485-618-2	Sequence 2, Appl
38	156	10.2	1161	1 US-08-485-618-99	Sequence 99, Appl
39	156	10.2	1161	1 US-08-362-652-2	Sequence 2, Appl
40	156	10.2	1161	2 US-08-605-672-2	Sequence 2, Appl
41	156	10.2	1161	2 US-08-605-672-99	Sequence 99, Appl
42	156	10.2	1161	2 US-08-482-293A-2	Sequence 2, Appl
43	156	10.2	1161	2 US-08-482-293A-99	Sequence 99, Appl
44	156	10.2	1161	2 US-08-943-363-2	Sequence 2, Appl
45	156	10.2	1161	2 US-08-943-363-99	Sequence 99, Appl

## ALIGNMENTS

RESULT 1  
US-08-286-889-46  
; Sequence 46, Application US/08286889  
; Patent No. 5470953  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Mich  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,889  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: F38,659  
REFERENCE/DOCKET NUMBER: 27866/32168  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-286-889-46

Query Match 10.9%; Score 166; DB 1; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
QY 10 PACYG-GFDLYFIIDKGSV-LHHWNEIYFVEQLAHKFTSPQURMSFIVSTRGTTLMK 67

Db 144 PEGGQEMDIAFLIDSGSIDOSDFTQMKDFVAKLMGQLASTSTSFSLMOYSNLKTHFT 203  
Qy 68 LTEDR-----EQIROGLEELQKVLPGDPTMHGEPERASEQIYYENROGYRTA-SVIAL 121  
Db 204 FTEFKSSLSPOSIVDAIVQ-----GLTYTASGIQKVVKLPFSKNGARKSAKKILIVI 258  
Qy 122 TDGELHEDLFFYSE--REANRSRDIGAIVCVGVKD-FNE-TOLARI-----ADSKOHVF 172  
Db 259 TDGQKFRDPLEYRHVYIPAEKA---GIRYAIQVGDAPREFPTALQELNTIGSAPSDHVF 315  
Qy 173 PVNDGFOLQGIHSLILKSCIEILAAEPSTICAGESFQVNVVNGGFRHARNVD 226  
Db 316 KVGK-FVALRSIQROIQEK---IFALCTESRSSSSFHQEMSGEGFSALSMD 364

## RESULT 2

US-08-485-618-46  
; Sequence 46, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-485-618-46

Query Match 10.9%; Score 166; DB 1; Length 1155;

Best Local Similarity 28.2%; Pred. No. 6e-09; Mismatches 96; Indels 30; Gaps 12;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 10 PACYG-GFDLYFILDSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTRGTLMK 67  
Db 144 PEGGQEMDIAFLIDSGSIDOSDFTQMKDFVAKLMGQLASTSTSFSLMOYSNLKTHFT 203

Qy 68 LTEDR-----EQIROGLEELQKVLPGDPTMHGEPERASEQIYYENROGYRTA-SVIAL 121  
Db 204 FTEFKSSLSPOSIVDAIVQ-----GLTYTASGIQKVVKLPFSKNGARKSAKKILIVI 258  
Qy 122 TDGELHEDLFFYSE--REANRSRDIGAIVCVGVKD-FNE-TOLARI-----ADSKOHVF 172  
Db 259 TDGQKFRDPLEYRHVYIPAEKA---GIRYAIQVGDAPREFPTALQELNTIGSAPSDHVF 315  
Qy 173 PVNDGFOLQGIHSLILKSCIEILAAEPSTICAGESFQVNVVNGGFRHARNVD 226  
Db 316 KVGK-FVALRSIQROIQEK---IFALCTESRSSSSFHQEMSGEGFSALSMD 364

## RESULT 3

US-08-362-652-46  
; Sequence 46, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-362-652-46

Query Match 10.9%; Score 166; DB 1; Length 1155;

Best Local Similarity 28.2%; Pred. No. 6e-09; Mismatches 96; Indels 30; Gaps 12;

Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

Qy 10 PACYG-GFDLYFILDSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTRGTLMK 67

Db 144 PEGGQEMDIAFLIDSGSIDOSDFTQMKDFVAKLMGQLASTSTSFSLMOYSNLKTHFT 203

Qy 68 LTEDR-----EQIROGLEELQKVLPGDPTMHGEPERASEQIYYENROGYRTA-SVIAL 121

Db 204 FTEFKSSLSPOSIVDAIVQ-----GLTYTASGIQKVVKLPFSKNGARKSAKKILIVI 258

Qy 122 TDGELHEDLFFYSE--REANRSRDIGAIVCVGVKD-FNE-TOLARI-----ADSKOHVF 172



Db 259 TDGKFRDPLEYRHVIPAERKA---GIIRYAIVGDAFREPTALQELNTIGSAPSODHVF 315  
Qy 173 PVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 226  
Db 316 KVGN-FVALNLSIQIOIEK----IFAIEGTESRSSSFQHEMSQEGFSALSM 364

## RESULT 4

US-08-605-672-46  
Sequence 46, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-46

Query Match 10.9%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 10 PACYG-GFDLYFLIDKSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTRTGLTMK 67  
Db 144 PECCGQEMDIAFLIDGSGSIDQSDPTQMKDFVALKMGQLASTSTSPSLMQSILNLTHT 203  
Qy 68 LTEDR-----EQIRQGLEIQLKVLPGSDTYMHGFEFASQIYYENRQGYRTA-SVIAL 121  
Db 204 FTEFKSLSPQSLVDIAIVQI-----GLTYASGIQKVVELPHSKNGAKSAKILIVI 258  
Qy 122 TDGSLHEDLFFYSE--REARSRDLGAIVYCVGYKD-FNE-TOLARI-----ADSKDHVF 172  
Db 259 TDGKFRDPLEYRHVIPAERKA---GIIRYAIVGDAFREPTALQELNTIGSAPSODHVF 315

Qy 173 PVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 226  
Db 316 KVGN-FVALNLSIQIOIEK----IFAIEGTESRSSSFQHEMSQEGFSALSM 364

## RESULT 5

US-08-482-293A-46  
Sequence 46, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-46

Query Match 10.9%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;  
Qy 10 PACYG-GFDLYFLIDKSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTRTGLTMK 67  
Db 144 PECCGQEMDIAFLIDGSGSIDQSDPTQMKDFVALKMGQLASTSTSPSLMQSILNLTHT 203  
Qy 68 LTEDR-----EQIRQGLEIQLKVLPGSDTYMHGFEFASQIYYENRQGYRTA-SVIAL 121  
Db 204 FTEFKSLSPQSLVDIAIVQI-----GLTYASGIQKVVELPHSKNGAKSAKILIVI 258  
Qy 122 TDGSLHEDLFFYSE--REARSRDLGAIVYCVGYKD-FNE-TOLARI-----ADSKDHVF 172  
Db 259 TDGKFRDPLEYRHVIPAERKA---GIIRYAIVGDAFREPTALQELNTIGSAPSODHVF 315  
Qy 173 PVNDGFOALQGIHSILKSCIEILAAEPSTICAGESFQVVRGNGFRHARNVD 226

Db 316 KVGK-FVALRSTQROIEK-----IFAIEGTSSRSSSSSFQHEMSQEGFSSALSMD 364

RESULT 6  
US-08-943-363-46

; Sequence 46, Application US/08943363  
; Patent No. 5837478  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Seair Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/943,363  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-943-363-46

Query Match 10.9%; Score 166; DB 2; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GFDLYFLDKSGSV-LHHMNEIYFVQOLAHKFIISPOLRMSFIVSTGTTLMK 67  
DB 144 PECEQEMDIAFLIDSGSIDSDPTQMKDFVAKLMGOLASTSTSPSLMOSNLKTHFT 203  
QY 68 LTEDR-----EQIRQGLEBLQKVLPGGDTYMEGFERASEQIYYENRQGYRTA-SVIAL 121  
DB 204 FTEFKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVELFHSKNGARSAKILLIIVI 258  
QY 122 TDGEIHEBLFFYSE--REANRSRDLGAIYVCVQKD-FNE-TQLARI-----ADSKDHVF 172  
DB 259 TDGQKFRDPLEYRHVYIPEAKA---GIRYAIQVGDARFETPLQELNLTIGSAPSODHVF 315  
QY 173 PVNDGFQALQGIHSLIKKSCIEILAAEPSTICAGESFQVYVVGNGFRHARNVD 226  
DB 316 KVGK-FVALRSTQROIEK-----IFAIEGTSSRSSSSSFQHEMSQEGFSSALSMD 364

RESULT 7  
US-09-193-043-46

; Sequence 46, Application US/09193043  
; Patent No. 6251395  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, Michael W.  
; TITLE OF INVENTION: No. 6251395el Human 2  
; FILE REFERENCE: 27866/35004  
; CURRENT APPLICATION NUMBER: US/09/193,043  
; CURRENT FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: 08/173,497  
; EARLIER FILING DATE: 1993-12-23  
; EARLIER APPLICATION NUMBER: 08/286,889  
; EARLIER FILING DATE: 1994-08-05  
; EARLIER APPLICATION NUMBER: 08/362,652  
; EARLIER FILING DATE: 1994-12-21  
; EARLIER APPLICATION NUMBER: 08/943,363  
; EARLIER FILING DATE: 1997-10-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 1155  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; US-09-193-043-46

Query Match 10.9%; Score 166; DB 3; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GFDLYFLDKSGSV-LHHMNEIYFVQOLAHKFIISPOLRMSFIVSTGTTLMK 67  
DB 144 PECEQEMDIAFLIDSGSIDSDPTQMKDFVAKLMGOLASTSTSPSLMOSNLKTHFT 203  
QY 68 LTEDR-----EQIRQGLEBLQKVLPGGDTYMEGFERASEQIYYENRQGYRTA-SVIAL 121  
DB 204 FTEFKSSLSPOSIVDAIVQLQ-----GLTYTASGIQKVVELFHSKNGARSAKILLIIVI 258  
QY 122 TDGEIHEBLFFYSE--REANRSRDLGAIYVCVQKD-FNE-TQLARI-----ADSKDHVF 172  
DB 259 TDGQKFRDPLEYRHVYIPEAKA---GIRYAIQVGDARFETPLQELNLTIGSAPSODHVF 315  
QY 173 PVNDGFQALQGIHSLIKKSCIEILAAEPSTICAGESFQVYVVGNGFRHARNVD 226  
DB 316 KVGK-FVALRSTQROIEK-----IFAIEGTSSRSSSSSFQHEMSQEGFSSALSMD 364

RESULT 8  
US-09-688-307A-46  
; Sequence 46, Application US/09688307A  
; Patent No. 6432404  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, Michael W.  
; TITLE OF INVENTION: No. 6432404el Human Beta-2  
; FILE REFERENCE: 27866/36646  
; CURRENT APPLICATION NUMBER: US/09/688,307A  
; CURRENT FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 09/193,043  
; PRIOR FILING DATE: 1998-11-16  
; PRIOR APPLICATION NUMBER: 08/605,672  
; PRIOR FILING DATE: 1996-02-22  
; PRIOR APPLICATION NUMBER: 08/173,497  
; PRIOR FILING DATE: 1993-12-23  
; PRIOR APPLICATION NUMBER: 08/286,889  
; PRIOR FILING DATE: 1994-08-05  
; PRIOR APPLICATION NUMBER: 08/362,652  
; PRIOR FILING DATE: 1994-12-21  
; PRIOR APPLICATION NUMBER: 08/943,363  
; PRIOR FILING DATE: 1997-10-03

NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-688-307A-46

Query Match 10.9%; Score 166; DB 4; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GPDLYFLDKSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTGTTLMK 67  
DB 144 PECGQEMDIAPFLIDGSGSIDQSDPTQMKDFVKALMGQLASTSTSFSLMQYSNLKTHFT 203  
QY 68 LTEDR-----EQIRQGLEBELQKVLPGSDTYMHGEPERASEQIYYENRQGYRTA-SVIAL 121  
DB 204 FTBKSSLSPOSVDALVQLQ-----GLTYTASGIQKVVLELPHSKNGARKSAKKILIVI 258  
QY 122 TDGELHEDLPFYSE--REARSRDLGAIYVCVGYD-FNE-TOLARI-----ADSKDHVF 172  
DB 259 TDGQKFRDPLEYRHVYIPEAKA---GIIRYAIQGDAREPTALQELNTTIGSAPSQDHVF 315  
QY 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRNGFRHARNVD 226  
DB 316 KVGK-FVALRSIQRIQIEK-----IFAIRGTSRSSSSFQHMMSGGFSALSMD 364

RESULT 9  
US-09-350-259-46

Sequence 46, Application US/09350259  
Patent No. 6620915  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 6620915el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/350,259  
CURRENT FILING DATE: 1999-07-08  
EARLIER APPLICATION NUMBER: 09/193,043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 46  
LENGTH: 1155  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-350-259-46

Query Match 10.9%; Score 166; DB 4; Length 1155;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GPDLYFLDKSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTGTTLMK 67  
DB 144 PECGQEMDIAPFLIDGSGSIDQSDPTQMKDFVKALMGQLASTSTSFSLMQYSNLKTHFT 203  
QY 68 LTEDR-----EQIRQGLEBELQKVLPGSDTYMHGEPERASEQIYYENRQGYRTA-SVIAL 121  
DB 204 FTBKSSLSPOSVDALVQLQ-----GLTYTASGIQKVVLELPHSKNGARKSAKKILIVI 258  
QY 122 TDGELHEDLPFYSE--REARSRDLGAIYVCVGYD-FNE-TOLARI-----ADSKDHVF 172  
DB 259 TDGQKFRDPLEYRHVYIPEAKA---GIIRYAIQGDAREPTALQELNTTIGSAPSQDHVF 315

QY 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRNGFRHARNVD 226  
DB 316 KVGK-FVALRSIQRIQIEK-----IFAIRGTSRSSSSFQHMMSGGFSALSMD 364

RESULT 10  
US-08-485-618-53

Sequence 53, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vieren, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-53

Query Match 10.9%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GPDLYFLDKSGSV-LHHMNEIYFVEQLAHKFIISPOLRMSFIVSTGTTLMK 67  
DB 144 PECGQEMDIAPFLIDGSGSIDQSDPTQMKDFVKALMGQLASTSTSFSLMQYSNLKTHFT 203  
QY 68 LTEDR-----EQIRQGLEBELQKVLPGSDTYMHGEPERASEQIYYENRQGYRTA-SVIAL 121  
DB 204 FTBKSSLSPOSVDALVQLQ-----GLTYTASGIQKVVLELPHSKNGARKSAKKILIVI 258  
QY 122 TDGELHEDLPFYSE--REARSRDLGAIYVCVGYD-FNE-TOLARI-----ADSKDHVF 172  
DB 259 TDGQKFRDPLEYRHVYIPEAKA---GIIRYAIQGDAREPTALQELNTTIGSAPSQDHVF 315  
QY 173 PVNDGFOALQGIHSLKSCIEILAAEPSTICAGESFOVVVRNGFRHARNVD 226

Db 316 KVGN-FVALRSIQROIQEK-----IFAIGTESRSSSSFOHEMSQEGFSSALSMD 364

## RESULT 11

US-08-362-652-53  
; Sequence 53, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-53

Query Match 10.9%; Score 166; DB 1; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GFDLYFLIDSGSV-LHMHNEIYYFVEQLAHKFPISPOLRMSFIVSTRGTTLMK 67  
DB 144 PECPQEMDIAPFLIDSGSIDQSDFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203  
QY 68 LTEDR-----EQIRGLEBELQKVLPGSDTYMHGFERASQEIYYENRQGYRTA-SVIAL 121  
DB 204 FTEFKSLSPQSLVDIAIVQD-----GLTYTAGIQLKVVLELPHSKNGARKSAKKILIVI 258  
QY 122 TDGELHEDLFFYSE--REANRSRDLAGIYVCVQKD-FNE-TOLARI-----ADSKDHF 172  
DB 259 TDQKFRDPLEYRHHVIPLEAKA---GIRYALIGVDAREPRLALQELINTIGSAPSODHVF 315  
QY 173 PVNDGFQALQGIHSLIKKSCIEIIAASPSTICAGESFOVVVNGGFRHARNVD 226  
DB 316 KVGN-FVALRSIQROIQEK-----IFAIGTESRSSSSFOHEMSQEGFSSALSMD 364

RESULT 12

US-08-605-672-53  
; Sequence 53, Application US/08605672  
; Patent No. 5817515  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/605,672  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32684  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1161 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-605-672-53

Query Match 10.9%; Score 166; DB 2; Length 1161;  
Best Local Similarity 28.2%; Pred. No. 6e-09;  
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACYG-GFDLYFLIDSGSV-LHMHNEIYYFVEQLAHKFPISPOLRMSFIVSTRGTTLMK 67  
DB 144 PECPQEMDIAPFLIDSGSIDQSDFTQMKDFVAKLMQGLASTSTSFSLMQYSNLKTHFT 203  
QY 68 LTEDR-----EQIRGLEBELQKVLPGSDTYMHGFERASQEIYYENRQGYRTA-SVIAL 121  
DB 204 FTEFKSLSPQSLVDIAIVQD-----GLTYTAGIQLKVVLELPHSKNGARKSAKKILIVI 258  
QY 122 TDGELHEDLFFYSE--REANRSRDLAGIYVCVQKD-FNE-TOLARI-----ADSKDHF 172  
DB 259 TDQKFRDPLEYRHHVIPLEAKA---GIRYALIGVDAREPRLALQELINTIGSAPSODHVF 315  
QY 173 PVNDGFQALQGIHSLIKKSCIEIIAASPSTICAGESFOVVVNGGFRHARNVD 226  
DB 316 KVGN-FVALRSIQROIQEK-----IFAIGTESRSSSSFOHEMSQEGFSSALSMD 364

RESULT 13  
US-08-482-293A-53  
; Sequence 53, Application US/08482293A

```
Patent No. 5831029
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vlieten, Monica
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,293A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-293A-53

Query Match 10.9%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 6e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACVG-GFDLYFLDKSGSV-LHHMNEIYVFEQLAHKFTSPQLRMSFIYVSTRTGLMK 67
DB 144 PEPCGQEMDLAFILDSGSDIDQSDFTQMKDFVAKMGQLASTSTSLMGIYSLIKTHFT 203
QY 68 LTEDR-----EQIRGLELQKVLPGSDTYMHGFEFASQIYENRQGYRTA-SVIAL 121
DB 204 FTEFKSLSPQSLVDAIVQLQ-----GLTYTASGIQKVKELPHSKNGAKRSKILIVI 258
QY 122 TDGELHDLFFYSE--REANRSLDAIYVCVYKD-FNE-TOLARI-----ADSKDHF 172
DB 259 TDGQKRPDLPEYHVIPEAKA---GIIRYALIGVGAFFRPTALQELINTIGSAPSDHVF 315
QY 173 PVNDGFOALGIIHSILKSCIEILAEPTICAGSFQVVVNGGFRHARNVD 226
DB 316 KVGN-FVALKSLIORIOEK---IFALGTESSRSSSFQHEMSQEGFSALSMD 364

RESULT 14
US-08-943-363-53
Sequence 53. Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Van der Vlieten, Monica
```

```
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vlieten, Monica
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-53

Query Match 10.9%; Score 166; DB 2; Length 1161;
Best Local Similarity 28.2%; Pred. No. 6e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

QY 10 PACVG-GFDLYFLDKSGSV-LHHMNEIYVFEQLAHKFTSPQLRMSFIYVSTRTGLMK 67
DB 144 PEPCGQEMDLAFILDSGSDIDQSDFTQMKDFVAKMGQLASTSTSLMGIYSLIKTHFT 203
QY 68 LTEDR-----EQIRGLELQKVLPGSDTYMHGFEFASQIYENRQGYRTA-SVIAL 121
DB 204 FTEFKSLSPQSLVDAIVQLQ-----GLTYTASGIQKVKELPHSKNGAKRSKILIVI 258
QY 122 TDGELHDLFFYSE--REANRSLDAIYVCVYKD-FNE-TOLARI-----ADSKDHF 172
DB 259 TDGQKRPDLPEYHVIPEAKA---GIIRYALIGVGAFFRPTALQELINTIGSAPSDHVF 315
QY 173 PVNDGFOALGIIHSILKSCIEILAEPTICAGSFQVVVNGGFRHARNVD 226
DB 316 KVGN-FVALKSLIORIOEK---IFALGTESSRSSSFQHEMSQEGFSALSMD 364

RESULT 15
US-09-193-043-53
Sequence 53. Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vlieten, Monica
```

```

; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 53
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-193-043-53

```

```

Query Match      10 9%; Score 166; DB 3; Length 1161;
Best Local Similarity 28.2%; Pred. No. 6e-09;
Matches 66; Conservative 42; Mismatches 96; Indels 30; Gaps 12;

```

```

Oy 10 PACYG-GPDLYFIIDKSGSV-LHHMNEIYYFVEQLAHKFI9POLRMSFIVSTRGTTLMK 67
Db 144 PECFQGEWDIAFLIDSGSIDQSDPTQMKDFKALMGQLASTSTSFSLMOYSNLIKTHFT 203
Oy 68 LTERD-----EQIQGLELELQKVLPGDPTVMHEGFERASEQIYYENROGYRTA-SVITIAL 121
Db 204 FTEFKSSISPSQSLVDALIVQLQ-----GLTYTASGIQKVKELFHSKNGARKSAKKILIVI 258
Oy 122 TDGELHEDLFFYSE--REANRSROLDGATVCGVVD-FNE--TOLARI-----ADSKDHVF 172
Db 259 TDGQKFRDPLEYRHVIFPAEKA--GIRYALGVGDAPREPTALQELNTTIGSAPSQDHVF 315
Oy 173 PVNDGFQALQGIHSHILKKSCEILIAEPSTICAGESFOVVVRGNGFRHARNVD 226
Db 316 KVGK-FVALRSIQKQIQEK-----IFALRGTSRSSSSSFQHEMSQEGFSSALSMQ 364

```

Search completed: June 13, 2005, 20:05:08  
 Job time : 29.4879 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 13, 2005, 20:03:29 ; Search time 93.9469 Seconds  
(without alignments)  
1195.537 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_28\_320

Perfect score: 1526  
Sequence: 1 OGGREDEGACGCGFDLYF.....DGLSFSSVITTTTCSGDG 293

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 1710399 seqs, 38334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1526	100.0	403	11 US-09-833-245-621
2	1526	100.0	551	14 US-10-038-307-18
3	1526	100.0	551	14 US-10-201-292-18
4	1526	100.0	564	10 US-09-918-715-187
5	1526	100.0	564	10 US-09-918-715-232
6	1526	100.0	564	14 US-10-301-822-199
7	1526	100.0	564	16 US-10-408-765A-183
8	1526	100.0	564	16 US-10-474-794-187
9	1526	100.0	564	16 US-10-474-794-232
10	1515	99.3	403	11 US-09-833-245-620
11	1514	99.2	333	10 US-09-796-753-12

12	1514	99.2	333	14 US-10-038-307-2	Sequence 2, App1
13	1514	99.2	333	14 US-10-201-292-2	Sequence 2, App1
14	1514	99.2	345	14 US-10-038-307-24	Sequence 24, App1
15	1514	99.2	345	14 US-10-201-292-24	Sequence 24, App1
16	1514	99.2	564	14 US-10-038-307-20	Sequence 20, App1
17	1514	99.2	564	14 US-10-201-292-20	Sequence 20, App1
18	1509	98.9	562	10 US-09-918-715-194	Sequence 194, App
19	1509	98.9	562	10 US-09-918-715-301	Sequence 301, App
20	1509	98.9	562	16 US-10-474-794-194	Sequence 194, App
21	1509	98.9	562	16 US-10-474-794-301	Sequence 301, App
22	1501	98.4	328	14 US-10-038-307-26	Sequence 26, App1
23	1501	98.4	328	14 US-10-201-292-26	Sequence 26, App1
24	1499.5	98.3	342	14 US-10-038-307-22	Sequence 22, App1
25	1499.5	98.3	342	14 US-10-201-292-22	Sequence 22, App1
26	1493	97.8	543	14 US-10-038-307-14	Sequence 14, App1
27	1493	97.8	543	14 US-10-038-307-16	Sequence 16, App1
28	1493	97.8	543	14 US-10-201-292-14	Sequence 14, App1
29	1493	97.8	543	14 US-10-201-292-16	Sequence 16, App1
30	1483	97.2	543	14 US-10-038-307-10	Sequence 10, App1
31	1483	97.2	543	14 US-10-201-292-10	Sequence 10, App1
32	1478	96.9	534	14 US-10-038-307-12	Sequence 12, App1
33	1478	96.9	534	14 US-10-201-292-12	Sequence 12, App1
34	1413	92.6	529	14 US-10-201-292-36	Sequence 36, App1
35	1299.5	85.2	504	14 US-10-201-292-34	Sequence 34, App1
36	1172	76.8	479	14 US-10-201-292-32	Sequence 32, App1
37	1058	69.3	460	14 US-10-201-292-38	Sequence 38, App1
38	1048	68.7	460	14 US-10-201-292-30	Sequence 30, App1
39	903	59.2	538	13 US-10-047-542-99	Sequence 99, App1
40	786	51.5	488	14 US-09-796-753-52	Sequence 52, App1
41	786	51.5	488	14 US-10-038-307-6	Sequence 6, App1
42	786	51.5	488	14 US-10-201-292-6	Sequence 6, App1
43	786	51.5	488	14 US-10-368-087-16	Sequence 16, App1
44	786	51.5	488	15 US-10-104-047-2639	Sequence 2639, App
45	773.5	50.7	587	9 US-09-764-870-312	Sequence 312, App

#### ALIGNMENTS

RESULT 1  
US-09-833-245-621  
Sequence 621, Application US/09833245  
Publication No. US20040010134A1  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Albumin Fusion Proteins  
FILE REFERENCE: PFS46PCT  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: 60/229, 358  
PRIOR FILING DATE: 2000-04-12  
PRIOR APPLICATION NUMBER: 60/256, 931  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/199, 384  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 2267  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 621  
LENGTH: 403  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-833-245-621

Query Match 100.0%; Score 1526; DB 11; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.1e-147;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 OGGREDEGACGCGFDLYFILDKSGSVLHHNMEIYVFQQLAHKFTSPQLRMSFIYFST 60  
Db 28 OGGREDEGACGCGFDLYFILDKSGSVLHHNMEIYVFQQLAHKFTSPQLRMSFIYFST 87  
Qy 61 RGTTLMTLTDREQIRGLEELQVLPDGSPTVMEGERASEQIYYENRQYRTASVITA 120

Db 88 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 147  
Qy 121 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 180  
Db 148 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 320

## RESULT 2

US-10-038-307-18  
Sequence 18, Application US/10038307  
Publication No. US20030134786A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKANNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038.307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 551  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-18

Query Match 100.0%; Score 1526; DB 14; Length 551;  
Best Local Similarity 100.0%; Pred. No. 1.8e-147;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 60  
Db 28 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 87  
Qy 61 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 120  
Db 88 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 147  
Qy 121 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 180  
Db 148 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 320

## RESULT 3

US-10-201-292-18  
Sequence 18, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKANNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292

Qy 1 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 60  
Db 28 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 87  
Qy 61 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 120  
Db 88 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 147  
Qy 121 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 180  
Db 148 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTTCAGSFQVVRGNGFRRANVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKMAALQVSMNDGLSFSSVITTTTHCSG 320

## RESULT 4

US-09-918-715-187  
Sequence 187, Application US/09918715  
Publication No. US20030017157A1  
GENERAL INFORMATION:  
APPLICANT: Bert Vogelstein  
APPLICANT: Brad St. Croix  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00134  
CURRENT APPLICATION NUMBER: US/09/918.715  
CURRENT FILING DATE: 2001-08-01  
PRIOR APPLICATION NUMBER: 60/222,599  
PRIOR FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: 60/224,360  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/282,850  
PRIOR FILING DATE: 2000-04-11  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 187  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-918-715-187

Query Match 100.0%; Score 1526; DB 10; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.9e-147;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 60  
Db 28 OGGRRBDGPGACYGFDLYFILDKSGSVLHNMNIYYFVQLAHKFISPOLRMSFIYFST 87  
Qy 61 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 120  
Db 88 RGTTLMLKLTEDREQIRQGLELQKVLPGGDTYMHGFERASEQIYYENRQGYRTASV11A 147  
Qy 121 LTDELHEDLFFYSEREANRSRDIGAIYVCVGDYFNETQIARIADSKDHVFPVNDGFOA 180



Db 148 LTDELHEDLFFYSEERANRSDLGAIYCVGVDFNETOLARIADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 320

## RESULT 5

US-09-918-715-232  
Sequence 232, Application US/09918715  
Publication No. US20030017157A1  
GENERAL INFORMATION:  
APPLICANT: Brad St. Croix  
APPLICANT: Bert Vogelstein  
APPLICANT: Kenneth Kinzler  
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS  
FILE REFERENCE: 1107.00134  
CURRENT FILING DATE: 2001-08-01  
PRIOR FILING DATE: 2001-08-01  
PRIOR APPLICATION NUMBER: 60/222,599  
PRIOR FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: 60/224,360  
PRIOR FILING DATE: 2000-08-11  
PRIOR APPLICATION NUMBER: 60/282,850  
PRIOR FILING DATE: 2000-04-11  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 232  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-918-715-232

Query Match 100.0%; Score 1526; DB 10; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.9e-147;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVVEQLAKHFTISPOLRMSFIVFST 60  
Db 28 QGGRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVVEQLAKHFTISPOLRMSFIVFST 87  
Qy 61 RGTLMKLTEDREOIRQGLELQKVLPGDITYMHGFERASEQIYENRQGYRTASVILIA 120  
Db 88 RGTLMKLTEDREOIRQGLELQKVLPGDITYMHGFERASEQIYENRQGYRTASVILIA 147  
Qy 121 LTDELHEDLFFYSEERANRSDLGAIYCVGVDFNETOLARIADSKDHVPVNDGFOA 180  
Db 148 LTDELHEDLFFYSEERANRSDLGAIYCVGVDFNETOLARIADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 320

## RESULT 6

US-10-301-822-199  
Sequence 199, Application US/10301822  
Publication No. US20030148410A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Schlegel, Robert

APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
APPLICANT: Burgart, Lawrence J.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF COLON CANCER  
FILE REFERENCE: MEM01-02392RNM  
CURRENT FILING DATE: US/10/301,822  
PRIOR FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 60/339,971  
PRIOR FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: US 60/361,978  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/381,988  
PRIOR FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 199  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-301-822-199

Query Match 100.0%; Score 1526; DB 14; Length 564;  
Best Local Similarity 100.0%; Pred. No. 1.9e-147;  
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVVEQLAKHFTISPOLRMSFIVFST 60  
Db 28 QGGRREDGPGACGCGFDLYFLIDKSGSVLHHMNEIYFVVEQLAKHFTISPOLRMSFIVFST 87  
Qy 61 RGTLMKLTEDREOIRQGLELQKVLPGDITYMHGFERASEQIYENRQGYRTASVILIA 120  
Db 88 RGTLMKLTEDREOIRQGLELQKVLPGDITYMHGFERASEQIYENRQGYRTASVILIA 147  
Qy 121 LTDELHEDLFFYSEERANRSDLGAIYCVGVDFNETOLARIADSKDHVPVNDGFOA 180  
Db 148 LTDELHEDLFFYSEERANRSDLGAIYCVGVDFNETOLARIADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 240  
Db 208 LOGIHSILKKSCEIILAAEPSTICAGSFQVYVNGNGFRHARVNDVLCSEFKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 293  
Db 268 NEKPSVEDTYLLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCSDG 320

## RESULT 7

US-10-408-765A-1823  
Sequence 1823, Application US/10408765A  
Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, Eoin D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Glenn, Gary M.  
APPLICANT: Warnock, Dale E.  
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
FILE REFERENCE: 660088.465  
CURRENT FILING DATE: US/10/408,765A  
PRIOR FILING DATE: 2003-04-04  
NUMBER OF SEQ ID NOS: 3077  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1823  
LENGTH: 564  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-408-765A-1823

	Query Match	100.0%;	Score 1526;	DB 16;	Length 564;
	Best Local Similarity	100.0%;	Pred. No. 1.9e-147;		
	Matches 293;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	QGGRRDEGPA	CGGFDLYFLIDKSGSVLHMNEIYYEVEOLAHFISPOLRMSFIVST	60	
Db	28	QGGRRDEGPA	CGGFDLYFLIDKSGSVLHMNEIYYEVEOLAHFISPOLRMSFIVST	87	
QY	61	RGTLMLKLTEDREQIRQGLBELQKYLPGQDTYMHGFFRASBOIYENRQGRITASVIL	120		
Db	88	RGTLMLKLTEDREQIRQGLBELQKYLPGQDTYMHGFFRASBOIYENRQGRITASVIL	147		
QY	121	LTDDELHEDLFYFEEEREARNSRDGAIYYCVGVDPNFTOLARIADSDNHYPVNDGFOA	180		
Db	148	LTDDELHEDLFYFEEEREARNSRDGAIYYCVGVDPNFTOLARIADSDNHYPVNDGFOA	207		
QY	181	LQGIHSHLTKKSCIEIILAAEPSTICAGESFOVYVRGNFRHARNDRLVCSFKINDSTVL	240		
Db	208	LQGIHSHLTKKSCIEIILAAEPSTICAGESFOVYVRGNFRHARNDRLVCSFKINDSTVL	267		
QY	241	NEKPSVEDTYLTCAPAPILKEVGMKAAIQVSMNDGLSFISSSVITTTTHCSDG	293		
Db	268	NEKPSVEDTYLTCAPAPILKEVGMKAAIQVSMNDGLSFISSSVITTTTHCSDG	320		

```

RESULT 8
US-10-474-794-187
; Sequence 187, Application US/10474794
; Publication No. US20040213793A1
; GENERAL INFORMATION:
; APPLICANT: Carlson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: US/10/474,794
; CURRENT FILING DATE: 2003-10-14
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 564
; TYPE: PRF
; ORGANISM: Homo sapiens
; US-10-474-794-187

```

	Query Match	100.0%;	Score 1526;	DB 16;	Length 564;
	Best Local Similarity	100.0%;	Pred. No. 1.9e+147;		
	Matches	293;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	QGGRRDDGPAICYGCGFDLYFTLDKSGSYLHHMNNLYYVEQLANKFISPOLRMSTFVST	60		
Db	28	QGGRRDDGPAICYGCGFDLYFTLDKSGSYLHHMNNLYYVEQLANKFISPOLRMSTFVST	87		
QY	61	RGTLMLKLTEDREQIQGLBELQKYLPGDGYMHGFFERASEQIYENRQGRATASVIA	120		
Db	88	RGTLMLKLTEDREQIQGLBELQKYLPGDGYMHGFFERASEQIYENRQGRATASVIA	147		
QY	121	LTDSELHEDLFFYSERENRNRSDLGALYYCYGVDFNTQTALRLADSNDHFFPVNDGQA	180		
Db	148	LTDSELHEDLFFYSERENRNRSDLGALYYCYGVDFNTQTALRLADSNDHFFPVNDGQA	207		
QY	181	LOGIHSILKSKSCIEILAAEPSTICAGSFQVVRNGNFRHARNDVRLCSFKINDSVTL	240		
Db	208	LOGIHSILKSKSCIEILAAEPSTICAGSFQVVRNGNFRHARNDVRLCSFKINDSVTL	267		
QY	241	NEKPSVEDVTLCPAPILKEVGMKAAALQVSMNDGLSTISSVITTTTHGCDG	293		
Db	268	NEKPSVEDVTLCPAPILKEVGMKAAALQVSMNDGLSTISSVITTTTHGCDG	320		

```

RESULT 9
US-10-474-794-232
; Sequence 232, Application US/10474794
; Publication No. US20040213793A1
GENERAL INFORMATION:
APPLICANT: Carson-Walter, Eleanor
APPLICANT: St. Croix, Brad
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth
TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
FILE REFERENCE: 1107, 00179
CURRENT APPLICATION NUMBER: US/10/474,794
CURRENT FILING DATE: 2003-10-14
PRIOR APPLICATION NUMBER: 60/282,850
PRIOR FILING DATE: 2001-04-11
PRIOR APPLICATION NUMBER: 60/309,829
PRIOR FILING DATE: 2001-08-01
NUMBER OF SEQ ID NOS: 359
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 232
; LENGTH: 564
; TYPE: PRY
; ORGANISM: Homo sapiens
US-10-474-794-232

```

	Query Match	Similarity	Score	DB	Length
Best Match	100.0%	100.0%	1526	DB 16	564
Matches	293	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
QY	1	QGGRRREGGACACYGFDLYFLIDKSGSVLHMNMNIYYFVEQLAHKFISPOLRMSEIVST	60		
Db	28	QGGRRREGGACACYGFDLYFLIDKSGSVLHMNMNIYYFVEQLAHKFISPOLRMSEIVST	87		
QY	61	RGTTLMKLTEDREQIROGLEELQVLPQGGDTYMHGFFERASEQIYYENRQGRITASVILIA	120		
Db	88	RGTTLMKLTEDREQIROGLEELQVLPQGGDTYMHGFFERASEQIYYENRQGRITASVILIA	147		
QY	121	LTTDELEHEDLFFYSERERANSRDGLAIVYCYGVQDQFNETQLARLADSDKHYPVNVDDGQA	180		
Db	148	LTTDELEHEDLFFYSERERANSRDGLAIVYCYGVQDQFNETQLARLADSDKHYPVNVDDGQA	207		
QY	161	LQGIHSLIKKSCIEIIAAEPSTICAGESFQVVYVRNGGFRHARVDRLCSPKINDSVTL	240		
Db	208	LQGIHSLIKKSCIEIIAAEPSTICAGESFQVVYVRNGGFRHARVDRLCSPKINDSVTL	267		
QY	241	NEKPFVEDTYLCPAPILKEVGKAKALQVSMNDGLSFRISSVITTTTHCSDG	293		
Db	268	NEKPFVEDTYLCPAPILKEVGKAKALQVSMNDGLSFRISSVITTTTHCSDG	320		

```

10 RESULT 10
11 US-09-833-245-620
12 ; Sequence 620, Application US/09833245
13 ; Publication No. US20040010134A1
14 ;
15 GENERAL INFORMATION:
16 ;
17 APPLICANT: Human Genome Sciences, Inc.
18 ;
19 TITLE OF INVENTION: Albumin Fusion Proteins
20 ;
21 FILE REFERENCE: PF546PCT
22 ;
23 CURRENT APPLICATION NUMBER: US/09/833,245
24 ;
25 PRIORITY FILING DATE: 2001-04-12
26 ;
27 PRIORITY APPLICATION NUMBER: 60/2229, 358
28 ;
29 PRIORITY FILING DATE: 2000-04-12
30 ;
31 PRIORITY APPLICATION NUMBER: 60/2256, 931
32 ;
33 PRIORITY FILING DATE: 2000-12-21
34 ;
35 PRIORITY APPLICATION NUMBER: 60/199, 384
36 ;
37 NUMBER OF SEQ ID NOS: 2267
38 ;
39 SOFTWARE: PatentIn Ver. 2.1
40 ;
41 SEQ ID NO 620
42 ;
43 LENGTH: 403
44 ;
45 TYPE: PRT
46 ;

```

```

/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (175)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (320)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (331)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (368)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-620

Query Match          99.3%; Score 1515; DB 11; Length 403;
Best Local Similarity 99.7%; Pred. No. 1,5e-146;
Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QGGRREDGPACGCGFDLYFLDKSGSVLHMNEIYFVEQLAHKFTSPQLRMSFIVFST 60
Db 28 QGGRREDGPACGCGFDLYFLDKSGSVLHMNEIYFVEQLAHKFTSPQLRMSFIVFST 87
Qy 61 RGTLMKLTEDREIQRLGLELQVLPFGDPTMHGGERASEQIYYENRQGYRTASVITA 120
Db 88 RGTLMKLTEDREIQRLGLELQVLPFGDPTMHGGERASEQIYYENRQGYRTASVITA 147
Qy 121 LTDELHEDLFFYSEERANRSDIGALVYCVGVDPNETOLARIADSKDHFPPVNDGFOA 180
Db 148 LTDELHEDLFFYSEERANRSDIGALVYCVGVDPNETOLARIADSKDHFPPVNDGFOA 207
Qy 181 LOGIHSILKSCIEIILAEPSTICAGSFQVVVNGGFRHARNVDRVLCSEFKINDSVTL 240
Db 208 LOGIHSILKSCIEIILAEPSTICAGSFQVVVNGGFRHARNVDRVLCSEFKINDSVTL 267
Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSSVIITTHCS 292
Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSSVIITTHCS 319

RESULT 11
US-09-796-753-12
/ Sequence 12, Application US/09796753
/ Publication No. US2003002798A1
/ GENERAL INFORMATION:
/ APPLICANT: McCarthy, Sean A.
/ TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
/ FILE REFERENCE: 7853-227-999
/ CURRENT APPLICATION NUMBER: US/09/796,753
/ PRIOR FILING DATE: 2001-03-01
/ PRIOR APPLICATION NUMBER: 09/183,175
/ PRIOR FILING DATE: 1998-10-30
/ PRIOR APPLICATION NUMBER: 09/223,094
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/223,546
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/224,246
/ PRIOR FILING DATE: 1998-12-30
/ PRIOR APPLICATION NUMBER: 09/259,388
/ PRIOR FILING DATE: 1999-02-26
/ PRIOR APPLICATION NUMBER: 60/122,458
/ PRIOR FILING DATE: 1999-03-01
/ PRIOR APPLICATION NUMBER: 09/312,359
/ PRIOR FILING DATE: 1999-05-14
/ PRIOR APPLICATION NUMBER: 09/336,536
/ PRIOR FILING DATE: 1999-06-18
/ PRIOR APPLICATION NUMBER: 09/342,687
/ PRIOR FILING DATE: 1999-06-29
/ PRIOR APPLICATION NUMBER: 09/345,464
/ PRIOR FILING DATE: 1999-06-30
/ PRIOR APPLICATION NUMBER: 09/365,164
/ PRIOR FILING DATE: 1999-07-30
```

```

/ PRIOR APPLICATION NUMBER: 09/399,723
/ PRIOR FILING DATE: 1999-09-20
/ PRIOR APPLICATION NUMBER: 09/409,634
/ PRIOR FILING DATE: 1999-09-30
/ PRIOR APPLICATION NUMBER: 09/471,179
/ PRIOR FILING DATE: 1999-12-23
/ PRIOR APPLICATION NUMBER: 09/474,071
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/474,072
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/514,010
/ PRIOR FILING DATE: 2000-02-25
/ PRIOR APPLICATION NUMBER: 09/516,745
/ PRIOR FILING DATE: 2000-03-01
/ PRIOR APPLICATION NUMBER: 09/572,002
/ PRIOR FILING DATE: 2000-05-14
/ PRIOR APPLICATION NUMBER: 09/597,993
/ PRIOR FILING DATE: 2000-06-19
/ PRIOR APPLICATION NUMBER: 09/599,596
/ PRIOR FILING DATE: 2000-06-22
/ PRIOR APPLICATION NUMBER: 09/630,334
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 09/606,565
/ PRIOR FILING DATE: 2000-06-29
/ PRIOR APPLICATION NUMBER: 09/606,317
/ PRIOR FILING DATE: 2000-06-29
/ PRIOR APPLICATION NUMBER: 09/665,666
/ PRIOR FILING DATE: 2000-09-20
/ PRIOR APPLICATION NUMBER: 09/677,751
/ PRIOR FILING DATE: 2000-09-30
/ NUMBER OF SEQ ID NOS: 162
/ SEQ ID NO 12
/ LENGTH: 333
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-796-753-12

Query Match          99.2%; Score 1514; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 1,5e-146;
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPACGCGFDLYFLDKSGSVLHMNEIYFVEQLAHKFTSPQLRMSFIVFST 60
Db 28 QGGRREDGPACGCGFDLYFLDKSGSVLHMNEIYFVEQLAHKFTSPQLRMSFIVFST 87
Qy 61 RGTLMKLTEDREIQRLGLELQVLPFGDPTMHGGERASEQIYYENRQGYRTASVITA 120
Db 88 RGTLMKLTEDREIQRLGLELQVLPFGDPTMHGGERASEQIYYENRQGYRTASVITA 147
Qy 121 LTDELHEDLFFYSEERANRSDIGALVYCVGVDPNETOLARIADSKDHFPPVNDGFOA 180
Db 148 LTDELHEDLFFYSEERANRSDIGALVYCVGVDPNETOLARIADSKDHFPPVNDGFOA 207
Qy 181 LOGIHSILKSCIEIILAEPSTICAGSFQVVVNGGFRHARNVDRVLCSEFKINDSVTL 240
Db 208 LOGIHSILKSCIEIILAEPSTICAGSFQVVVNGGFRHARNVDRVLCSEFKINDSVTL 267
Qy 241 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSSVIITTHCS 291
Db 268 NEKPSVEDTYLLCPAPILKEVGKALQVSMNDGLSFSSSVIITTHCS 318

RESULT 12
US-10-018-307-2
/ Sequence 2, Application US/10038307
/ Publication No. US20030134786A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKANAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
```

;; CURRENT APPLICATION NUMBER: US/10/038.307  
;; CURRENT FILING DATE: 2002-06-28  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 333  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-038-307-2

Query Match 99.2%; Score 1514; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1,5e-146;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 60  
Db 28 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 87  
Qy 61 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 120  
Db 88 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 147  
Qy 121 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 180  
Db 148 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 240  
Db 208 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 291  
Db 268 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 318

RESULT 13  
US-10-201-292-2

;; Sequence 2, Application US/10201292  
;; Publication No. US20030144193A1  
;; GENERAL INFORMATION:  
;; APPLICANT: James B. ROTTMAN  
;; APPLICANT: Theresa L. O'KEEFE  
;; APPLICANT: Engin OZKAYNAK  
;; APPLICANT: Judith J. HEALEY  
;; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
;; FILE REFERENCE: 7853-253-999  
;; CURRENT APPLICATION NUMBER: US/10/201,292  
;; CURRENT FILING DATE: 2003-02-14  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 2  
;; LENGTH: 333  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-201-292-2

Query Match 99.2%; Score 1514; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1,5e-146;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 60  
Db 28 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 87  
Qy 61 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 120  
Db 88 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 147  
Qy 121 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 180  
Db 148 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 240

Db 208 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 291  
Db 268 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 318

RESULT 14  
US-10-038-307-24

;; Sequence 24, Application US/10038307  
;; Publication No. US20030134786A1  
;; GENERAL INFORMATION:  
;; APPLICANT: James B. ROTTMAN  
;; APPLICANT: Theresa L. O'KEEFE  
;; APPLICANT: Engin OZKAYNAK  
;; APPLICANT: Judith J. HEALEY  
;; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
;; FILE REFERENCE: 7853-253-999  
;; CURRENT APPLICATION NUMBER: US/10/038.307  
;; CURRENT FILING DATE: 2002-06-28  
;; NUMBER OF SEQ ID NOS: 26  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 24  
;; LENGTH: 345  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-038-307-24

Query Match 99.2%; Score 1514; DB 14; Length 345;  
Best Local Similarity 100.0%; Pred. No. 1,6e-146;  
Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 60  
Db 28 QGGRREDGPGACGFDLYFLIDKSGSVLHNMNEIYFVEQLAHKFISPOLRMSFIVST 87  
Qy 61 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 120  
Db 88 RGTTLMKLTEDREQIRQGLELQKVLPGDPTMHGFERASEQIYENRQGYRTASVITA 147  
Qy 121 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 180  
Db 148 LTDELHEDLFFYSERANRSDGAIYVCVGDENFETOLARLADSKDHVPVNDGFOA 207  
Qy 181 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 240  
Db 208 LOGIHSILKSCIEILAAEPSTICAGESFOVVVNGNGFRHARVDRVLCSPKINDSVTL 267  
Qy 241 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 291  
Db 268 NEKPSVEDTYLLCPAPILKEVGKAKALQVSMNDGLSFISSVIITTHCS 318

RESULT 15  
US-10-201-292-24

;; Sequence 24, Application US/10201292  
;; Publication No. US20030144193A1  
;; GENERAL INFORMATION:  
;; APPLICANT: James B. ROTTMAN  
;; APPLICANT: Theresa L. O'KEEFE  
;; APPLICANT: Engin OZKAYNAK  
;; APPLICANT: Judith J. HEALEY  
;; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
;; FILE REFERENCE: 7853-253-999  
;; CURRENT APPLICATION NUMBER: US/10/201,292  
;; CURRENT FILING DATE: 2003-02-14  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 24  
;; LENGTH: 345  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens

US-10-201-292-24

Query Match 99.2%; Score 1514; DB 14; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-146;  
 Matches 291; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 QGGRREDGGPACVCGFDLYFLIDKSGSVLHHMNEIYYFVEQLAHKFTSPQLRMSFIVEST 60
DB 28 QGGRREDGGPACVCGFDLYFLIDKSGSVLHHMNEIYYFVEQLAHKFTSPQLRMSFIVEST 87

QY 61 RGTTLMLKLTEDREQIROGLBELQKVLPGGDTYMHGFEFRASEQIYENRQGYRTASVYIA 120
DB 88 RGTTLMLKLTEDREQIROGLBELQKVLPGGDTYMHGFEFRASEQIYENRQGYRTASVYIA 147

QY 121 LTDGELHEDLFFYGEREARNRSRDIGAIYVCVKDFNETQLARIADSKDHVFPVNDGFQA 180
DB 148 LTDGELHEDLFFYGEREARNRSRDIGAIYVCVKDFNETQLARIADSKDHVFPVNDGFQA 207

QY 181 LQGIHSHILKKSCEIILAEPSTICAGESFQVYVRGNGFRHARVNDVYLCSEKINDSVTL 240
DB 208 LQGIHSHILKKSCEIILAEPSTICAGESFQVYVRGNGFRHARVNDVYLCSEKINDSVTL 267

QY 241 NEKPFVYEDTYLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCS 291
DB 268 NEKPFVYEDTYLCPAPILKEVGKKAALQVSMNDGLSFISSVIITTHCS 318
    
```

Search completed: June 13, 2005, 20:36:50  
 Job time : 94.9469 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2005, 19:47:08 ; Search time 21.0213 Seconds

(without alignments)

1341.095 Million cell updates/sec

Title: us-09-970-076-2\_copy\_28\_320

Perfect score: 1526

Sequence: 1 QGGRREDGPGACGCGFDLYF.....DGLSFSSSVITTTTCSGDG 293

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	9.5	1153	1 RWHU1B	cell surface glyco
2	143.5	9.4	1163	1 RWHU1C	cell surface glyco
3	143	9.4	3124	2 A40020	collagen alpha 1(X
4	141	9.2	1153	2 S00551	leukocyte surface
5	139	9.1	724	2 A48569	antigen Em100 - B1
6	137	9.0	1747	2 A45974	collagen alpha 1(X
7	134.5	8.8	712	2 A45638	immunodominant mic
8	131	8.6	760	1 C2MS	classical-compleme
9	130	8.5	1857	2 S31212	collagen alpha 1(X
10	130	8.5	1888	2 S78476	collagen alpha 1(X
11	128	8.4	3051	2 S42373	hypothetical prote
12	118	7.7	929	2 I51027	type XII collagen
13	113.5	7.4	496	2 A37979	cartilage matrix p
14	113.5	7.4	764	1 BBHU	complement factor
15	111.5	7.3	2944	2 A54849	collagen alpha 1(V
16	110.5	7.2	761	1 BBMS	complement factor
17	109.5	7.2	1029	1 S21369	collagen alpha 2(V
18	108.5	7.1	500	2 S66522	cartilage matrix p
19	107.5	7.0	493	2 A33809	cartilage matrix p
20	105.5	6.9	567	2 T28797	hypothetical prote
21	103	6.7	747	2 I51579	complement factor
22	100.5	6.6	574	2 A46283	sporocyste surface
23	100	6.6	460	2 T23087	hypothetical prote
24	99.5	6.5	597	2 S33578	top protein - fru1
25	99	6.5	559	2 S04531	thrombospondin-rel
26	99	6.5	698	2 B96958	dnax protein (heat
27	99	6.5	817	2 T50240	kinesin-like prote
28	99	6.5	917	2 S09646	collagen alpha 2(V
29	99	6.5	1018	1 CGHU2A	collagen alpha 2(V

30	98.5	6.5	537	2 T04822	hypothetical prote
31	98	6.4	741	2 T46488	hypothetical prote
32	98	6.4	3176	2 CGHU3A	collagen alpha 3(V
33	97.5	6.4	642	2 H81185	dnax protein NMB05
34	97	6.4	3137	2 A37797	collagen alpha 3(V
35	96.5	6.3	676	2 T47637	hypothetical prote
36	96	6.3	272	2 A55346	integrin alpha-1 -
37	96	6.3	689	2 F84811	probable retrovira
38	95.5	6.3	334	2 AF1166	transcription regu
39	95.5	6.3	932	2 JCS953	inter-alpha-inhibi
40	95.5	6.3	1179	2 A53213	integrin alpha-E c
41	94.5	6.2	1151	2 A45226	integrin alpha-1 c
42	94	6.2	843	2 A40970	undulin 1 - human
43	94	6.1	642	2 E70121	hypothetical prote
44	92.5	6.1	642	2 B81917	probable chaperone
45	92	6.0	292	2 B83736	transposase (04) B

## ALIGNMENTS

### RESULT 1

RWHU1B

cell surface glycoprotein CD11b precursor [validated] - human

N/Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Ma

eukocyte integrin alpha chain; neutrophil adherence receptor alpha chain

C/Species: Homo sapiens (man)

C/Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 09-Jul-2004

C/Accession: A31108; A26915; A41600; A30892; A32218; A46526; A26091; I52567

R/Corbi: A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.

J. Biol. Chem. 263, 12403-12411, 1988

A/Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD

B.

A/Reference number: A31108; MUID:88315033; PMID:2457584

A/Accession: A31108

A/Molecule type: mRNA

A/Residues: 1-1153 <COR>

A/Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148

A/Note: part of this sequence was confirmed by protein sequencing

R/Arnaut, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.

J. Cell Biol. 106, 2153-2158, 1988

A/Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M

A/Reference number: A28915; MUID:88257215; PMID:2454931

A/Accession: A28915

A/Molecule type: mRNA

A/Residues: 1-499, 501-965, 'P', 967-1153 <ARN>

A/Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594

A/Note: the authors translated the codon TAC for residue 1129 as Thr

A/Note: part of this sequence, including the amino end of the mature protein, was confir

R/Shelley, C.S.; Arnaut, M.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991

A/Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regu

A/Reference number: A41600; MUID:92073318; PMID:1683702

A/Accession: A41600

A/Molecule type: DNA

A/Residues: 1-9 <SHR>

A/Cross-references: GB:M76724; NID:G180018; PIDN:AAA5410.1; PID:G553215

R/Arnaut, M.A.; Remold-O'Donnell, B.; Pierce, M.W.; Harris, P.; Tenen, D.G.

A/Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion

A/Reference number: A94193; MUID:88190151; PMID:2853753

A/Accession: A30892

A/Molecule type: mRNA

A/Residues: 917-1042 <AR2>

A/Cross-references: GB:M18044

R/Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.

Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989

A/Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor

A/Reference number: A32218; MUID:89098893; PMID:2563162

A/Accession: A32218

A/Molecule type: mRNA

A/Residues: 9-1153 <HIC>

A/Cross-references: GB:U04145; NID:G189068; PIDN:AAA59903.1; PID:G386975

N.Alternate names: leukocyte adhesion receptor p150,95 alpha chain  
C.Species: Homo sapiens (man)  
C.Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text\_change 09-Jul-2004

R.Corbidi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 12750-12751, 1990  
A.Reference number: A36584  
A.Contents: extratum  
A.Accession: A36584  
A.Molecule type: DNA  
A.Residues: 1-1163 <COR>  
A.Cross-references: UNIPROT:P20702  
A.Note: This revision to the sequence from reference A35543 includes the carboxyl end  
R.Corbidi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2788, 1990  
A.Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A.Reference number: A35543; MUID:90153906; PMID:2303426  
A.Accession: A35543  
A.Molecule type: DNA  
A.Residues: 1-834 <CO2>  
A>Note: This sequence has been revised in reference A36584  
R.Corbidi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
EMBO J. 6, 4023-4028, 1987  
A>Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A.Reference number: S00864; MUID:88166645; PMID:3327667  
A.Accession: S00864  
A.Molecule type: mRNA  
A.Residues: 1-755,'L','757'-1163 <CO3>  
A.Cross-references: GB:M81695; EMBL:Y00093; NID:g487829; PIDN:AAA59180.1; PID:g487830  
A.Note: part of this sequence was confirmed by protein sequencing  
C.Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on myeloid cells:  
C.Genetics:  
A.Gene: GDB:ITGAX; CD11c  
A.Cross-references: GDB:119758; OMIM:151510  
A.Map position: 16p11.2-16p11.2  
C.Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo,  
C.Keywords: calcium, cell adhesion, glycoprotein, heterodimer; magnesium; tandem repeat;  
F.II-19/Domain: signal sequence #status predicted <SIG>  
F.II-20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>  
F.III-1107/Domain: extracellular #status predicted <EXT>  
F.IV-319/Domain: von Willebrand factor type A repeat homology <VMA4>  
F.VIII-1133/Domain: transmembrane #status predicted <TM>  
F.XIII-1163/Domain: intracellular #status predicted <INT>  
F.F61.89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre-

Query Match            9.4%; Score 143.5; DB 1; Length 1163;  
Best Local Similarity   24.8%; Pred.No.0.0024;

	Matches	Conservative	%	Mismatches	Gaps	Indels
Oy	17	DIVFLDLSGSV-LHHMNEIYYFEVDLAHFKTISPOLRMSTVSTRGTLMKLTEDRE--	73	151	DIVFLDGSGSSSRNFATMFNVRAVISQFORPSQTFSLMQSNNKFOTHTFFEEFRRTS	210
Oy	74	--ORRGLEELOKVLPGDDTYMHGEFGFBASDEIYYENRCGYRAS-VIALTDGELLHD	129	211	NPLSLASYHQ-----GFYTATAIQVVVAHLRFNASYGARDATKLILVIITDSKEGD	265
Oy	130	IIFYSEREARRSDIGAIYVCVG-----KDFTETOLARIAD-SKHVPFYNDG	177	266	SLLDYMD-VIAPMADAAGIIIRYAIGVGLAFQRNSWKELD-----IASKPSEGHIFKYED-	318
Oy	178	FOTALGGIHSHLKKSCTIELAAEPSTTCAGESEQVAVVRNGCF	219	319	PDALDKDIOMQLTEK-----IFAIEGTETTSSSFELLEMAOEGF	356

RESULT 3  
A40020 collagen alpha 1(XII) chain precursor - chicken  
N.Alternate names: fibrochmerin  
C.Species: Gallus gallus (chicken)  
C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004  
C.Accession: A40020; A34485; B34485; A28037; S233814; S222524; S28811



R.Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shimomura, T.; Tanaka, H.; Nishida, Y.; Obay  
 U. Cell Biol. 115, 209-221, 1991  
 A>Title: The complete primary structure of type XII collagen shows a chimeric molecule w  
 nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.  
 A/Reference number: AA0020, MUID:92011862; PMID:1918137  
 A/Accession: AA0020  
 A/Molecule type: mRNA  
 A/Residues: 1-3124 <YAM>  
 A/Cross-references: UNIPROT:P13944; GB:D00824; NID:9222810; PIDN:BA00701.1; PID:9222811  
 A/Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and,  
 R.Gordon, M.K.; Gerecke, D.R.; Dublet, B.; van der Rest, M.; Olsen, B.R.  
 J. Biol. Chem. 264, 19772-19778, 1989  
 A>Title: Type XII collagen. A large multidomain molecule with partial homology to type I  
 A/Reference number: A34485; MUID:90062079; PMID:2584192  
 A/Accession: A34485  
 A/Molecule type: mRNA  
 A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>  
 A/Cross-references: EMBL:J05137; NID:G211284; PIDN:AAA48635.1; PID:G211285  
 A/Accession: B34485  
 A/Molecule type: Protein  
 A/Residues: 2772-2792;2846-2873 <GOR2>  
 R.Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987  
 A>Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c  
 A/Reference number: A28037; MUID:87317590; PMID:3476925  
 A/Accession: A28037  
 A/Molecule type: mRNA  
 A/Residues: 2960-2976, 'P', 2978-3074, 'AG' <GOR3>  
 A/Cross-references: EMBL:M17375; NID:G211649; PIDN:AAA48718.1; PID:G211650  
 A/Note: This sequence has been revised in reference A34485  
 R.Koch, M.; Bernasconi, C.; Chiquet, M.  
 Eur. J. Biochem. 207, 847-856, 1992  
 A>Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of  
 A/Reference number: S23814; MUID:92362621; PMID:1323460  
 A/Accession: S23814  
 A/Molecule type: Protein  
 A/Residues: 'X', 1333, 'Q', 1335-1347;1394-1928;2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <  
 R.Dublet, B.; van der Rest, M.  
 J. Biol. Chem. 262, 17724-17727, 1987  
 A>Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peptin-  
 A/Reference number: S22254; MUID:88087065; PMID:3121603  
 A/Accession: S22254  
 A/Molecule type: Protein  
 A/Residues: 2831-2832, 'T', 2834, 'R', 2836-2843;3002-3014 <DUB>  
 R.Tueb, J.; Tueb, B.  
 Biochim. Biophys. Acta 1171, 97-98, 1992  
 A>Title: The two splice variants of collagen XII share a common 5' end.  
 A/Reference number: S28811; MUID:93042014; PMID:1420368  
 A/Accession: S28811  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-24,1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>  
 A/Cross-references: EMBL:X67327  
 C/Genetics:  
 A/Introns: 2845/3; 2863/3; 2887/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1  
 C/Keywords: alternative splicing; cell binding; coll binding; coll; connective tissue; disulfid  
 F/1-23/Domain: signal sequence #status predicted <SIG>  
 F/24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>  
 F/24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
 F/24-114/Domain: I11A #status predicted <I11A>  
 F/24-105/Domain: fibronectin type III repeat homology <FN3A>  
 F/137-301/Domain: von Willebrand factor type A repeat homology <VWA1>  
 F/133-425/Domain: I11B #status predicted <I11B>  
 F/133-414/Domain: fibronectin type III repeat homology <FN3B>  
 F/437-601/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F/629-1178/Domain: I11C #status predicted <I11C>  
 F/630-711/Domain: fibronectin type III repeat homology <FN3C>  
 F/721-802/Domain: fibronectin type III repeat homology <FN3D>  
 F/812-895/Domain: fibronectin type III repeat homology <FN3E>  
 F/905-986/Domain: fibronectin type III repeat homology <FN3F>  
 F/995-1076/Domain: fibronectin type III repeat homology <FN3G>  
 F/1086-1169/Domain: fibronectin type III repeat homology <FN3H>  
 F/1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>

F/1384-2295/Domain: I11D #status predicted <I11D>  
 F/1384-1465/Domain: fibronectin type III repeat homology <FN3I>  
 F/1474-1557/Domain: fibronectin type III repeat homology <FN3J>  
 F/1566-1647/Domain: fibronectin type III repeat homology <FN3K>  
 F/1655-1738/Domain: fibronectin type III repeat homology <FN3L>  
 F/1756-1838/Domain: fibronectin type III repeat homology <FN3M>  
 F/1847-1928/Domain: fibronectin type III repeat homology <FN3N>  
 F/1937-2019/Domain: fibronectin type III repeat homology <FN3O>  
 F/2018-2110/Domain: fibronectin type III repeat homology <FN3P>  
 F/2119-2199/Domain: fibronectin type III repeat homology <FN3Q>  
 F/2207-2299/Domain: fibronectin type III repeat homology <FN3R>  
 F/2325-2490/Domain: von Willebrand factor type A repeat homology <VWA4>  
 F/2438-2440/Region: cell adhesion #status predicted  
 F/2509-2750/Domain: IXP, homologous to NC4 domain of type IX collagen #status predicted  
 F/2751-2902/Domain: collagenous COL2 #status predicted <COL2>  
 F/2899-2901/Region: cell attachment (R-G-D) motif  
 F/2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>  
 F/2946-3048/Domain: collagenous COL1 #status predicted <COL1>  
 F/3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>  
 F/32,1006,1032,1044,1512,1767,2210,2273,2532,2683/Binding site: carbohydrate (Asn) (cova)  
 F/2780,2789,2836,2842,2860,2869,3004,3007/Modified site: hydroxyproline (Pro) #statu

Query Match 9.4%; Score 143; DB 2; Length 3124;

Best Local Similarity 26.1%; Pred. No. 0.0097; Matches 65; Conservative 51; Mismatches 97; Indels 36; Gaps 14;

17 DLYFLDKSGSV-LHHNNEIYFVEQLAHKF-ISP-QLRMSFIVSTRTTLMKLTEDRE 73

Db 439 DVEFLVGVSGSIGANVVKRAFLFVYKSFELSPKVOISLVQYKR--DPHHEFSLNRY 496

Qy 74 QIRGSELEKQVP--GGDTYMEGFERRASBQIYENRQGR--TASVITATDGEIHEH 129

Db 497 NRVVDIIQAIINTFPRYRGSGTNTGKAMTYVEKVFVTSK--GSRPVVPWMLITDCK--SSD 554

Qy 130 LFFSESEARNSDPLKAIYVCVKDPEFQRLARID--SKDHFPPNDGFQALQGIHS 187

Db 555 AF--KBPATKLRAVDYIFRAVGKDVKLELAIPAPETHYTED--FDMEQRISFE 610

Qy 188 ILKSCIEIILAEPRSTICAGESFOVVVRGNGFRHARV--DRYLCSFKINDSVTLNEKP 244

Db 611 LTQGVLCIRI-----EQELAIKRSKYPAPKAMVSDVTSDSFKWSKMSAGSE-- 657

Qy 245 FSVEDTYLL 253

Db 658 ---EKSYLI 663

#### RESULT 4

S00551 leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse

N/Alternate names: complement-3 receptor alpha chain

C/Species: Mus musculus (house mouse)

C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004

C/Accession: S00551; I59078

R/PyTela, R.

EMBO J. 7, 1371-1378, 1988

A>Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the ir

A/Reference number: S00551; MUID:88312584; PMID:3044779

A/Accession: S00551

A/Molecule type: DNA

A/Residues: 1-1153 <PYT>

A/Cross-references: UNIPROT:P05555; EMBL:X07640; NID:952982; PIDN:CAA30479.1; PID:952983

A/Note: the authors translated the codon CAC for residue 569 as Gln

R.Saetre, L.; Roman, J.M.; Teglow, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,

Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986

A>Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recept

A/Reference number: I59078; MUID:86287312; PMID:2942940

A/Accession: I59078

A/Molecule type: DNA

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Residues: 11-44 <RES>

A/Cross-references: GB:M14293; NID:g198993; PIDN:AAA39484.1; PID:9554193

C/Genetics:

A:Gene: Mac-1  
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
C:Keywords: cell adhesion; glycoprotein; transmembrane protein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental  
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 9.2%; Score 141; DB 2; Length 1153;

Best Local Similarity 24.2%; Pred. No. 0.0038;  
Matches 71; Conservative 50; Mismatches 96; Indels 76; Gaps 16;

Qy 17 DLYFLDKSGSVLHHNNEIYFVEQLAHKFIS---PQLRMSFIVSTRTGTTLMKLTED-- 71  
Db 150 DIVFLIDSGSIS---NNIDF---QKMKPFVSTWEGPKSKTLFS-----LMQYSDER 197  
Qy 72 ----REQIRQGLHELOKVLKLP---GGDTYMHGFERASEQIYIE--NRGQYTAIVIALT 122  
Db 198 IHFTFNDKRNKPSRSHSPITKQLNGRTKTASGIRKVRRELPHKTNGARENAAKILVIT 257  
Qy 123 DGELEHDLFPYSE--REANRSRDLAGIYVCVKD--FNETOLARIAD-----SKDHVP 173  
Db 258 DGEKRGDLDYKDVLPEDRA---GVIRYVGVGNAPFKPQSRRELDITAKRPHENFQ 314  
Qy 174 VNDGFQALQGIHSLTKSCIEILAAPESTCAGESFQVVVRNGFRRHNRVDRVLSFK 233  
Db 315 V-DNFEALNTIQNLQOEK---IFAIEGTGTGTSSTSFHEMSQEGF----- 355  
Qy 234 INDSVTLANEKPSVEDTYLCLCAPILKEVG---MKALQVSMNDGLSFISS 282  
Db 356 -SASITSN-----GPIILGSGVSPDMAGAFLYTSDKQVTFINTT 393

## RESULT 5

A48569  
Antigen Em100 - Elmeria maxima

C:Species: Elmeria maxima

C>Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C:Accession: A48569

R:Paranonten, L.; Hug, D.; Willebrand, M.; Weber, G.

Mol. Biochem. Parasitol. 57, 171-174, 1993

A:Title: Sequence of a major Elmeria maxima antigen homologous to the Elmeria tenella m

A:Reference number: A48569; MUID:93149203; PMID:8426611

A:Accession: A48569

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-724 <PAS>

A:CROSS-References: UNIPROT:004588; GB:M99058; NID:9158890; PID:9158891

A:Note: Sequence extracted from NCBI backbone (NCBIN:123776, NCBI:P:123777)

F:45-218/Domain: von Willebrand factor type A repeat homology <VWA2>

F:236-326/Domain: chromospondin type 1 repeat homology <THR1>

F:309-371/Domain: chromospondin type 1 repeat homology <THR2>

F:372-437/Domain: chromospondin type 1 repeat homology <THR3>

F:433-493/Domain: chromospondin type 1 repeat homology <THR4>

F:494-556/Domain: chromospondin type 1 repeat homology <THR5>

F:560-610/Domain: chromospondin type 1 repeat homology <THR6>

Query Match 9.1%; Score 139; DB 2; Length 724;

Best Local Similarity 26.6%; Pred. No. 0.003; Indels 34; Gaps 14;

Matches 59; Conservative 40; Mismatches 89; Indels 34; Gaps 14;

Qy 12 CYGFDPLVFLDKSGSV-LHHNNEIYFVEQLAHKF-ISP-QLRMSFIVSTRTGTTLMKL 68  
Db 42 CTRLLDVMVLVDESSIGTSIGSYNGKVRSFISNFAQTMPSPDDVRVGLVTFGSAVTRMDL 101  
Qy 69 TEDRQIRQGLHELOKVLKLP---GGDTYMHGFERASEQIYENRQGYR--TASVIALTDG 124  
Db 102 SDSRQNDMLAAAKKQPYAGSTYTHGLAKA--EELTFSPQGRNAPMMLVMTDG 160  
Qy 125 ELHEDLFPYSERE-----ANRSRDLAGIYVCVG--KQFNETOLARIA--DSKHV-PP- 173  
Db 161 A-----SSRSQTLAAEKLRNRGVIIIVLVGTVGNASRCSINAGCCTSDVECPR 212

Qy 174 -VNDGFQALQGIHSLTKSCIEI---LAAPESTI--CAGE 208  
Db 213 YLQSNCGVSSQINGIITKACDLADAVCSEWSEYGPCEG 254

## RESULT 6

A45974  
Collagen alpha 1(XIV) chain precursor, short form 2 - chicken

N:Alternate names: undulin

C:Species: Gallus gallus (chicken)

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C:Accession: A45974; S30085; S22916; S17035; S20833

R:Gercke, D.R.; Foley, J.W.; Castagnola, P.; Gemari, M.; Dublet, B.; Cancedda, R.; Lin

J. Biol. Chem. 268, 12177-12184, 1993

A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' regions

A:Reference number: A45974; MUID:93280195; PMID:8505337

A:Accession: A45974

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-1747 <GER>

A:CROSS-References: UNIPROT:P32018

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBI:P:133365)

R:Apfe, S.S.

submitted to the EMBL data library, March 1992

A:Reference number: S30085

A:Accession: S30085

A:Molecule type: mRNA

A:Residues: 1472-1660 <APT>

A:CROSS-References: EMBL:X65122; NID:962871; PID:CAA46238.1; PID:9938175

R:Trueb, J.; Trueb, B.

Eur. J. Biochem. 207, 549-557, 1992

A:Title: Type XIV collagen is a variant of undulin.

A:Reference number: S22916; MUID:92339443; PMID:1339349

A:Accession: S22916

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>

R:Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayr, T.F.; van der Rest, M.; Mayne, F

Eur. J. Biochem. 201, 333-338, 1991

A:Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens v

A:Reference number: S17035; MUID:92037585; PMID:1935930

A:Accession: S17035

A:Molecule type: mRNA

A:Residues: 1472-1659 <GOR1>

A:Accession: S20833

A:Molecule type: protein

A:Residues: 1551-1570, 1593-1599, 1639-1667 <GOR2>

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer

F:40-204/Domain: von Willebrand factor type A repeat homology <VWA1>

F:236-317/Domain: fibronectin type III repeat homology <FN3A>

F:326-409/Domain: fibronectin type III repeat homology <FN3B>

F:418-498/Domain: fibronectin type III repeat homology <FN3C>

F:507-591/Domain: fibronectin type III repeat homology <FN3D>

F:625-707/Domain: fibronectin type III repeat homology <FN3E>

F:716-798/Domain: fibronectin type III repeat homology <FN3F>

F:806-893/Domain: fibronectin type III repeat homology <FN3G>

F:924-1089/Domain: von Willebrand factor type A repeat homology <VWA2>

F:1111-1355/Domain: non-collagenous NC2 #status predicted <NC2>

F:1511-1553/Domain: non-collagenous NC2 #status predicted <NC2>

F:1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 9.0%; Score 137; DB 2; Length 1747;

Best Local Similarity 25.4%; Pred. No. 0.014; Indels 26; Gaps 12;

Matches 63; Conservative 45; Mismatches 114; Indels 26; Gaps 12;

Qy 17 DLYFLDKSGSV-LHHNNEIYFVEQL-AHKEFISPO-LRMSFIVSTRTGTTLMKLT--E 70  
Db 926 DIVFLVDSGWSIGDNNKILISFLYSTVGALDKIGPGTGVALIQFDDPRTBEKLANAYK 985  
Qy 71 DREQIRQGLHELOKVLKLP---GGDTYMHGFERASEQIYENRQGYR--TASVIALTDGELHE 128

Db 966 TKTETLEAIQOI--AYKAGNTTKTGAKIHKARE-VLFTGAGRRKCI PKVLVITTDGRSOD 1042  
Qy 129 DLFFYSEREANRSRL-GAIVCVGVKDFNETOLARIAD--SKDHVFPVNDGFQALQGI 185  
Db 1043 DV-----NKVREMGDLDFSPFALIVADADYSELVNISKSBERVFFVD--FDAPFTIE 1096  
Qy 186 HSLKSKCTEILAEAPSTTCAGESFOVVVRGNGFPHANNDRVLCSFKINDSVTLNEKPF 245  
Db 1097 DELTIFVETASATCPLVFKQDNFA-----GFKMEMFGILVEKFSALDGVSMEPGT 1150  
Qy 246 SVEDTYLL 253  
Db 1151 NVYPCYRL 1158  
RESULT 7  
A45638  
Immunodominant microneme protein Ebp100 - Eimeria tenella  
C/Species: Eimeria tenella  
C/Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A45638  
R/Tomley, F.M.; Clarke, L.E.; Kawasoe, U.; Djikema, R.; Kok, J.J.  
Mol. Biochem. Parasitol. 49, 277-288, 1991  
A/Title: Sequence of the gene encoding an immunodominant microneme protein of Eimeria tenella  
A/Reference number: A45638; MUID:92131064; PMID:1775171  
A/Accession: A45638  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-712 <TOM>  
A/Cross-references: UNIPROT:O43981; GB:AFO32905; GB:M73495; NID:92707732; PIDN:AAD03350.  
A/Note: Sequence extracted from NCBI backbone (NCBI:77752, NCBI:77756)  
F/148-216/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/238-286/Domain: chromospondin type 1 repeat homology <THR1>  
F/309-371/Domain: chromospondin type 1 repeat homology <THR2>  
F/372-432/Domain: chromospondin type 1 repeat homology <THR3>  
F/433-493/Domain: chromospondin type 1 repeat homology <THR4>  
F/494-556/Domain: chromospondin type 1 repeat homology <THR5>  
F/560-610/Domain: chromospondin type 1 repeat homology <THR6>  
Query Match 8.8%; Score 134.5; DB 2; Length 712;  
Best Local Similarity 24.6%; Pred. No. 0.0068;  
Matches 55; Conservative 40; Mismatches 88; Indels 41; Gaps 11;  
Qy 12 CYGGEFDLYFLDKSGSV-LHHMNEIYFVEQLAHKF-ISPQ-LHMSFVSTRTTLMKL 68  
Db 45 CTSLDLVWLVDGSGISGTSNFKVRCPIEDPVNSMPTSPEDVAVGLTPT----- 96  
Qy 69 TEDREQIROLGELQKVLPG-----GDTVMHSGFERASEQIYYENQGR-- 113  
Db 97 ---RSKYRMNLSDPKATPISLAISARSLSYSTGVTYTHYGLQDA-KKLVDYVNAARNN 152  
Qy 114 TASYIIALTOGELHEDLFFYSEREANRSRDIGAIYVCGV-KDRENOQLARIADSKOHVF 172  
Db 153 VPKVLVLTWTDAA--SNLPSQTRSSAAMALRDGAILVVLGVSGVNSSECRSIACSTSN 210  
Qy 173 P-----VNDGFQALQGIHSLKSCIEILAEAPSTTC--CAGE 208  
Db 211 PRYQSNMNSVNTQVNGIIRAKCDLAKDAVCSEMSSEYGPVGE 254  
RESULT 8  
C2MS  
classical-complement-pathway C3/C5 convertase (BC 3.4.21.43) C2 component precursor - mc  
N/Alternate names: C3 convertase; C5 convertase; complement C2  
C/Species: Mus musculus (house mouse)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: A38876; B36593; I54429  
R/Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.  
submitted to GenBank, January 1991  
A/Reference number: A38875  
A/Accession: A38876  
A/Molecule type: DNA

A/Residues: 1-760 <IS2>  
A/Cross-references: UNIPROT:P21180; GB:M57891; GB:J05661; NID:9192436; PIDN:AAA63294.1; R/Ishikawa, N.; Nonaka, M.; Wetsel, R.A.; Colten, H.R.  
J. Biol. Chem. 265, 19040-19046, 1990  
A/Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different mechanisms for cleavage of C2 and factor B  
A/Reference number: A36593; MUID:91035430; PMID:2229060  
A/Accession: B36593  
A/Molecule type: mRNA  
A/Residues: 1-760 <ISH>  
A/Cross-references: EMBL:M57891; NID:9192436; PIDN:AAA63294.1; PID:9192437  
R/Falut, A.; Wakeland, R.K.; McConnell, T.J.; Gitlin, D.; Whitehead, A.S.; Colten, H.R.  
Immunogenetics 25, 290-298, 1987  
A/Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.  
A/Reference number: I54429; MUID:87192938; PMID:2883115  
A/Accession: I54429  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 660-677, R', 679, 681-723, G', 725 <RES>  
A/Cross-references: GB:M16211; NID:9199289; PIDN:AAA39562.1; PID:9199290  
A/Genetics:  
A/Introns: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 414/1; 461/1; 492/3; 537/1  
C/Function: The proenzyme forms a complex with C4a and is activated by cleavage into C2a, C2b and C2c  
A/Description: cleaves complement C3 and complement C5 alpha chains  
A/Pathway: complement classical pathway  
C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; C2; C2b; C2c; C2d; C2e; C2f; C2g; C2h; C2i; C2j; C2k; C2l; C2m; C2n; C2o; C2p; C2q; C2r; C2s; C2t; C2u; C2v; C2w; C2x; C2y; C2z; C2aa; C2ab; C2ac; C2ad; C2ae; C2af; C2ag; C2ah; C2ai; C2aj; C2ak; C2al; C2am; C2an; C2ao; C2ap; C2aq; C2ar; C2as; C2at; C2au; C2av; C2aw; C2ax; C2ay; C2az; C2ba; C2bb; C2bc; C2bd; C2be; C2bf; C2bg; C2bh; C2bi; C2bj; C2bk; C2bl; C2bm; C2bn; C2bo; C2bp; C2bq; C2br; C2bs; C2bt; C2bu; C2bv; C2bw; C2bx; C2by; C2bz; C2ca; C2cb; C2cc; C2cd; C2ce; C2cf; C2cg; C2ch; C2ci; C2cj; C2ck; C2cl; C2cm; C2cn; C2co; C2cp; C2cq; C2cr; C2cs; C2ct; C2cu; C2cv; C2cw; C2cx; C2cy; C2cz; C2da; C2db; C2dc; C2dd; C2de; C2df; C2dg; C2dh; C2di; C2dj; C2dk; C2dl; C2dm; C2dn; C2do; C2dp; C2dq; C2dr; C2ds; C2dt; C2du; C2dv; C2dw; C2dx; C2dy; C2dz; C2ea; C2eb; C2ec; C2ed; C2ee; C2ef; C2eg; C2eh; C2ei; C2ej; C2ek; C2el; C2em; C2en; C2eo; C2ep; C2eq; C2er; C2es; C2et; C2eu; C2ev; C2ew; C2ex; C2ey; C2ez; C2fa; C2fb; C2fc; C2fd; C2fe; C2ff; C2fg; C2fh; C2fi; C2fj; C2fk; C2fl; C2fm; C2fn; C2fo; C2fp; C2fq; C2fr; C2fs; C2ft; C2fu; C2fv; C2fw; C2fx; C2fy; C2fz; C2ga; C2gb; C2gc; C2gd; C2ge; C2gf; C2gg; C2gh; C2gi; C2gj; C2gk; C2gl; C2gm; C2gn; C2go; C2gp; C2gq; C2gr; C2gs; C2gt; C2gu; C2gv; C2gw; C2gx; C2gy; C2gz; C2ha; C2hb; C2hc; C2hd; C2he; C2hf; C2hg; C2hi; C2hj; C2hk; C2hl; C2hm; C2hn; C2ho; C2hp; C2hq; C2hr; C2hs; C2ht; C2hu; C2hv; C2hw; C2hx; C2hy; C2hz; C2ia; C2ib; C2ic; C2id; C2ie; C2if; C2ig; C2ih; C2ii; C2ij; C2ik; C2il; C2im; C2in; C2io; C2ip; C2iq; C2ir; C2is; C2it; C2iu; C2iv; C2iv; C2iw; C2ix; C2iy; C2iz; C2ja; C2jb; C2jc; C2jd; C2je; C2jf; C2jg; C2jh; C2ji; C2jj; C2jk; C2jl; C2jm; C2jn; C2jo; C2jp; C2jq; C2jr; C2js; C2jt; C2ju; C2jv; C2jw; C2jx; C2jy; C2jz; C2ka; C2kb; C2kc; C2kd; C2ke; C2kf; C2kg; C2kh; C2ki; C2kj; C2kl; C2km; C2kn; C2ko; C2kp; C2kq; C2kr; C2ks; C2kt; C2ku; C2kv; C2kw; C2kx; C2ky; C2kz; C2la; C2lb; C2lc; C2ld; C2le; C2lf; C2lg; C2lh; C2li; C2lj; C2lk; C2ll; C2lm; C2ln; C2lo; C2lp; C2lq; C2lr; C2ls; C2lt; C2lu; C2lv; C2lw; C2lx; C2ly; C2lz; C2ma; C2mb; C2mc; C2md; C2me; C2mf; C2mg; C2mh; C2mi; C2mj; C2mk; C2ml; C2mn; C2mo; C2mp; C2mq; C2mr; C2ms; C2mt; C2mu; C2mv; C2mw; C2mx; C2my; C2mz; C2na; C2nb; C2nc; C2nd; C2ne; C2nf; C2ng; C2nh; C2ni; C2nj; C2nk; C2nl; C2nm; C2nn; C2no; C2np; C2nq; C2nr; C2ns; C2nt; C2nu; C2nv; C2nw; C2nx; C2ny; C2nz; C2oa; C2ob; C2oc; C2od; C2oe; C2of; C2og; C2oh; C2oi; C2oj; C2ok; C2ol; C2om; C2on; C2oo; C2op; C2oq; C2or; C2os; C2ot; C2ou; C2ov; C2ow; C2ox; C2oy; C2oz; C2pa; C2pb; C2pc; C2pd; C2pe; C2pf; C2pg; C2ph; C2pi; C2pj; C2pk; C2pl; C2pm; C2pn; C2po; C2pp; C2pq; C2pr; C2ps; C2pt; C2pu; C2pv; C2pw; C2px; C2py; C2pz; C2qa; C2qb; C2qc; C2qd; C2qe; C2qf; C2qg; C2qh; C2qi; C2qj; C2qk; C2ql; C2qm; C2qn; C2qo; C2qp; C2qq; C2qr; C2qs; C2qt; C2qu; C2qv; C2qw; C2qx; C2qy; C2qz; C2ra; C2rb; C2rc; C2rd; C2re; C2rf; C2rg; C2rh; C2ri; C2rj; C2rk; C2rl; C2rm; C2rn; C2ro; C2rp; C2rq; C2rr; C2rs; C2rt; C2ru; C2rv; C2rw; C2rx; C2ry; C2rz; C2sa; C2sb; C2sc; C2sd; C2se; C2sf; C2sg; C2sh; C2si; C2sj; C2sk; C2sl; C2sm; C2sn; C2so; C2sp; C2sq; C2sr; C2ss; C2st; C2su; C2sv; C2sw; C2sx; C2sy; C2sz; C2ta; C2tb; C2tc; C2td; C2te; C2tf; C2tg; C2th; C2ti; C2tj; C2tk; C2tl; C2tm; C2tn; C2to; C2tp; C2tq; C2tr; C2ts; C2tt; C2tu; C2tv; C2tw; C2tx; C2ty; C2tz; C2ua; C2ub; C2uc; C2ud; C2ue; C2uf; C2ug; C2uh; C2ui; C2uj; C2uk; C2ul; C2um; C2un; C2uo; C2up; C2uq; C2ur; C2us; C2ut; C2uu; C2uv; C2uw; C2ux; C2uy; C2uz; C2va; C2vb; C2vc; C2vd; C2ve; C2vf; C2vg; C2vh; C2vi; C2vj; C2vk; C2vl; C2vm; C2vn; C2vo; C2vp; C2vq; C2vr; C2vs; C2vt; C2vu; C2vv; C2vw; C2vx; C2vy; C2vz; C2wa; C2wb; C2wc; C2wd; C2we; C2wf; C2wg; C2wh; C2wi; C2wj; C2wk; C2wl; C2wm; C2wn; C2wo; C2wp; C2wq; C2wr; C2ws; C2wt; C2wu; C2wv; C2ww; C2wx; C2wy; C2wz; C2xa; C2xb; C2xc; C2xd; C2xe; C2xf; C2xg; C2xh; C2xi; C2xj; C2xk; C2xl; C2xm; C2xn; C2xo; C2xp; C2xq; C2xr; C2xs; C2xt; C2xu; C2xv; C2xw; C2xx; C2xy; C2xz; C2ya; C2yb; C2yc; C2yd; C2ye; C2yf; C2yg; C2yh; C2yi; C2yj; C2yk; C2yl; C2ym; C2yn; C2yo; C2yp; C2yq; C2yr; C2ys; C2yt; C2yu; C2yv; C2yw; C2yx; C2yy; C2yz; C2za; C2zb; C2zc; C2zd; C2ze; C2zf; C2zg; C2zh; C2zi; C2zj; C2zk; C2zl; C2zm; C2zn; C2zo; C2zp; C2zq; C2zr; C2zs; C2zt; C2zu; C2zv; C2zw; C2zx; C2zy; C2zz  
Query Match 8.6%; Score 131; DB 1; Length 760;  
Best Local Similarity 23.2%; Pred. No. 0.014;  
Matches 74; Conservative 56; Mismatches 115; Indels 74; Gaps 16;  
Qy 14 GGFQVLYTLKSSGVSLHNMNEY-YFVQOLAHKFIQPOLR--MSFYESTRGTTLMKLT 70  
Db 258 GHMLVYLLDASQSVTEKDFIFKSAELMVERIFSEFVAVTVAITFASOPTIMSLIS 317  
Qy 71 DREQIROLGELQKVLPGDPTVMHSGFERASEQIYYE-----NROGYRTAS- 116  
Db 318 ERSG-----DVEEVITSLSASAYKHENATGANTYEVLRVYSMMQTOCDRLGMETS 371  
Qy 117 -----VIALTDELHEDLFFYSEREANRSRDIGAI-----VYCVGV----- 155  
Db 372 KEIRHTIILITDCK--SNMGDSPEKAVTRIRRELIETQNRNDYLDIYAIIVGKLDVDMKE 429  
Qy 156 FNETQARIADSDXDPVVDGFOALQGIHSLKSCIEIILAEAPSTTCAGESFOVVVR 215  
Db 430 LNE--LSSKQDGRHHAITLDA--KALQQLFEHNLVDK-----TDITCG----- 472  
Qy 216 GNGFPHANVDRV--LCSFKINDSVTLNEKPSVSDTYLLCPAPILKEVGMK--AAQVDS 271  
Db 473 GNSANASDQERTPWQVTFKPKSKETQGS--LISQWVTLAAHCFRDIOMEDHMLKRVN 530  
Qy 272 MND-----GLSPISSVYI 285  
Db 531 VGDPTSGHKEFLVEDYII 549  
RESULT 9  
S3122  
collagen alpha 1(XIV) chain precursor, short form - chicken  
C/Species: Gallus gallus (chicken)

C>Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 15-Sep-2003  
C/Accession: S31212  
R:Maechli, C.; Trueb, U.; Kessler, B.; Winterhalter, K.H.; Trueb, B.  
Eur. J. Biochem. 212, 483-490, 1993  
A>Title: Complete primary structure of chicken collagen XIV.  
A/Reference number: S31211; MUID:93185668; PMID:8444186  
A/Accession: S31212  
A/Structure: nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-1857 <WAE>  
A/Cross-references: EMBL:X70792; NID:G288874; PIDN:CA50063.1; PID:G288875  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C/Genetics:  
A/Gene: COL14A1  
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>  
F:29-110/Domain: fibronectin type III repeat homology <FN3A>  
F:156-330/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:352-433/Domain: fibronectin type III repeat homology <FN3B>  
F:442-525/Domain: fibronectin type III repeat homology <FN3C>  
F:534-614/Domain: fibronectin type III repeat homology <FN3D>  
F:623-707/Domain: fibronectin type III repeat homology <FN3E>  
F:741-823/Domain: fibronectin type III repeat homology <FN3F>  
F:832-914/Domain: fibronectin type III repeat homology <FN3G>  
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>  
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1857;  
Best Local Similarity 25.0%; Pred. No. 0.055; Indels 26; Gaps 12;  
Matches 62; Conservative 44; Mismatches 116;  
Qy 17 DLVFLDKSGSV-LHHMNEIYFVEQL-AHKFISPO-LRMSFIVSTRGTLMKLT--E 70  
Db 1042 DLVFLVDSWMSIGDDNFKKISFLYSTVGALDKIGDGTQVAILQFSDDPRTFKLNAYK 1101  
Qy 71 DREQIRQGLEELQKVLPGDVTYMEGFERASEQIYYENRGYRTA--SVIALLDGELEH 128  
Db 1102 TKEITLLEAIQOI--AVKGGNTKTKGKAIKHARE-VLFTGEAGMKKGI PKVLVITDGRSQD 1158  
Qy 129 DLFPFSEBRANRDL-GAIVYCVGVKDPNFTQALRIAD--SKDHYFPVNDGQALQGIT 185  
Db 1159 DV-----NKVSRMQLDGSPFAIGVADADYSELVNISSKSPSRHVPFVD--FDAFTKIE 1212  
Qy 186 HSILKSCIEILAEPTICAGESFOVVVRGFRHARNDVRLCSFKINDSYTLNEKPF 245  
Db 1213 DELITPFCETASATCPVLVFDGDKLA-----GFMGMEMGVLEKFSALDGVSMERCTF 1266  
Qy 246 SVEDTYLL 253  
Db 1267 NVYPCYRL 1274

RESULT 10  
S78476  
collagen alpha 1(XIV) chain precursor, long form - chicken  
C/Species: Gallus gallus (chicken)  
C>Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 09-Jul-2004  
C/Accession: S78476; S31211  
R:Trueb, B.  
submitted to the EMBL Data Library, January 1993  
A/Reference number: S78476  
A/Accession: S78476  
A/Molecule type: mRNA  
A/Residues: 1-1888 <TRU>  
A/Cross-references: UNIPROT:P32018; EMBL:X70793; NID:G288872; PIDN:CA50064.1; PID:G2888  
R:Maechli, C.; Trueb, U.; Kessler, B.; Winterhalter, K.H.; Trueb, B.  
Eur. J. Biochem. 212, 483-490, 1993  
A>Title: Complete primary structure of chicken collagen XIV.  
A/Reference number: S31211; MUID:93185668; PMID:8444186  
A/Accession: S31211  
A/Structure: preliminary  
A/Molecule type: mRNA

A/Residues: 1-416;1460-1811,1843-1888 <WAE>  
A/Cross-references: EMBL:X70793  
C/Genetics:  
A/Gene: COL14A1  
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>  
F:29-110/Domain: fibronectin type III repeat homology <FN3A>  
F:156-330/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:352-433/Domain: fibronectin type III repeat homology <FN3B>  
F:442-525/Domain: fibronectin type III repeat homology <FN3C>  
F:534-614/Domain: fibronectin type III repeat homology <FN3D>  
F:623-707/Domain: fibronectin type III repeat homology <FN3E>  
F:741-823/Domain: fibronectin type III repeat homology <FN3F>  
F:832-914/Domain: fibronectin type III repeat homology <FN3G>  
F:922-1009/Domain: fibronectin type III repeat homology <FN3H>  
F:1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 8.5%; Score 130; DB 2; Length 1888;  
Best Local Similarity 25.0%; Pred. No. 0.057; Indels 26; Gaps 12;  
Matches 62; Conservative 44; Mismatches 116;  
Qy 17 DLVFLDKSGSV-LHHMNEIYFVEQL-AHKFISPO-LRMSFIVSTRGTLMKLT--E 70  
Db 1042 DLVFLVDSWMSIGDDNFKKISFLYSTVGALDKIGDGTQVAILQFSDDPRTFKLNAYK 1101  
Qy 71 DREQIRQGLEELQKVLPGDVTYMEGFERASEQIYYENRGYRTA--SVIALLDGELEH 128  
Db 1102 TKEITLLEAIQOI--AVKGGNTKTKGKAIKHARE-VLFTGEAGMKKGI PKVLVITDGRSQD 1158  
Qy 129 DLFPFSEBRANRDL-GAIVYCVGVKDPNFTQALRIAD--SKDHYFPVNDGQALQGIT 185  
Db 1159 DV-----NKVSRMQLDGSPFAIGVADADYSELVNISSKSPSRHVPFVD--FDAFTKIE 1212  
Qy 186 HSILKSCIEILAEPTICAGESFOVVVRGFRHARNDVRLCSFKINDSYTLNEKPF 245  
Db 1213 DELITPFCETASATCPVLVFDGDKLA-----GFMGMEMGVLEKFSALDGVSMERCTF 1266  
Qy 246 SVEDTYLL 253  
Db 1267 NVYPCYRL 1274

RESULT 11  
S42373  
hypothetical protein T20G5.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C>Date: 07-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-May-2004  
C/Accession: S42373  
R:Smith, A.  
submitted to the EMBL Data Library, March 1994  
A/Reference number: S42373  
A/Accession: S42373  
A/Molecule type: DNA  
A/Residues: 1-3051 <SMT>  
A/Cross-references: EMBL:Z30423; NID:G458479; PID:G458485  
C/Genetics:  
A/Insertions: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/2; 1041/1; 1146/1;  
F:512-679/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:754-793/Domain: fibronectin type II repeat homology <2F1>  
F:1201-1244/Domain: BGF homology <EGF>

Query Match 8.4%; Score 128; DB 2; Length 3051;  
Best Local Similarity 27.1%; Pred. No. 0.15; Indels 66; Gaps 13;  
Matches 64; Conservative 31; Mismatches 75;  
Qy 12 CYGGF-----DLVFLDKSGSVLHH--WNEIYFVEQLA 43  
Db 483 CYGFPDVSSNANLPPGRVCTVQTCERKQRTDLVFLDGSIGSYVFKNEVLPFAFEV 542  
Qy 44 HKFI--SPQLRMSFIVSTRGTLMKLT--DREQIRQGLEELQKVLPG-----GDTVMH- 94  
Db 543 ELFEIGSKTRVGLIQSDQIRHFPDLDQYGDSDLKIGISFTQ-YLTGLTRGALQHM 601

Qy 95 --EGFRASRQIYENRQGR-----TASVITATDGEHLDELFFYSREANRSDLAGI 147  
 Db 602 VQEGF-----SERRRARPQOSDIARVAIITLDGREQDQNV-----TGPADSAKRLSIN 648  
 Qy 148 VYCVGVKD-FNRTOLARIADSKDHPVYVDGFQALQGIHSITLKSCTEIIIAEFS 202  
 Db 649 TPAIVGTHVLASLESLISGSPNRWFFV-DKFKDLDTRLRSMIOK-----AACPS 697

## RESULT 12

type XII collagen alpha-1 chain - eastern newt (fragment)  
 C/Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
 C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
 C/Accession: 151027  
 R/Met, Y.; Yang, E.V.; Klatte, K.P.; Tassava, R.A.  
 Dev. Biol. 168, 503-513, 1995  
 A/Title: Monoclonal antibody MT2 identifies the urodele alpha 1 chain of type XII collagen  
 A/Reference number: 151027; MUID:95246925; PMID:7729585  
 A/Accession: 151027  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-929 <MBI>  
 A/Cross-references: UNIPROT:091145; EMBL:U9494; NID:g632647; PIDN:AAA80217.1; PID:g6326  
 F155-236/Domain: fibronectin type III repeat homology <3FR>  
 F1631-795/Domain: von Willebrand factor type A repeat homology <WMA3>

Query Match 7.7%; Score 118; DB 2; Length 929;  
 Best Local Similarity 21.2%; Pred. No. 0.21; Mismatches 135; Indels 26; Gaps 9;  
 Matches 57; Conservative 51;

Qy 17 DLVYILDKSGV-LHMNEIYFVEQLAHKFIISPOLRMSFTVSTRTGLTKLTEDRQI 75  
 Db 633 DIVLLVDSWMSIGRNFKIVRISRVVEVPDIGSDRQIVASQVSGDPRTEWQNTKTI 692  
 Qy 76 RQGLEELQKLP--GGDTYMHGFERASEQIYENRQGR-----RTASVITATDGEHL 129  
 Db 693 KKSIMDAVANLPYKGGNTNTSGALFIIENNF---RPGVGMREKARKIATILTDGKSQDD 749  
 Qy 130 LFFYSREANRSDGAIYCVGVDPFNETOLARIADSKDHPVNP-DGFOALQGIHSI 188  
 Db 750 IVASRKRVA---DEGILYAVGINADENELKEIASPPDELYMTNVADFSLTNTIVDL 805  
 Qy 189 LKKSCTEIIIAA---EPSTICAGESFOVVVNGFRRHARNVDRVLCSEKINDSVTLNEKP 244  
 Db 806 TENVCNSYKGGQGLNPNLVTSEPTPRSFRTWPPSGSYER----FKVEYYPVAGSRP 861  
 Qy 245 FSVEDTYLLCAPILKEVGMKALQVSN 273  
 Db 862 --QEVYVRGTQTTVLVLGLKRETEYYN 887

## RESULT 13

cartilage matrix protein precursor - human  
 C/Species: Homo sapiens (man)  
 C/Date: 12-Jul-1991 #sequence\_revision 12-Jul-1991 #text\_change 09-Jul-2004  
 C/Accession: A37979  
 R/Jenkins, R.N.; Osborne-Lawrence, S.L.; Sinclair, A.K.; Eddy Jr., R.L.; Byers, M.G.; Sh  
 J. Biol. Chem. 265, 19624-19631, 1990  
 A/Title: Structure and chromosomal location of the human gene encoding cartilage matrix  
 A/Reference number: A37979; MUID:91060568; PMID:2246248  
 A/Accession: A37979  
 A/Molecule type: DNA  
 A/Residues: 1-496 <JEN>  
 A/Cross-references: UNIPROT:P21941; GB:J05667  
 A/Accession: B37979  
 A/Molecule type: mRNA  
 A/Residues: 157-250, 'L', 292-496 <JEN>  
 A/Cross-references: GB:M55683; GB:J05666; GB:J05667; NID:g180651; PIDN:AAA63904.1; PID:g  
 C/Accession: A37979  
 A/Title: Complete primary structure for the zymogen of human complement factor B. Sequence of the

A/Cross-references: GDB:127280; OMIM:115437  
 A/Map position: 1p35-1p35  
 A/Introns: 32/1; 147/3; 222/1; 264/1; 403/1; 454/1; 481/1  
 C/Complex: homotrimer  
 C/Superfamily: cartilage matrix protein; EGF homology; von Willebrand factor type A repe  
 C/Keywords: glycoprotein, homotrimer  
 F1-42/Domain: signal sequence #status predicted <SIG>  
 F123-496/Product: cartilage matrix protein #status predicted <MAT>  
 F139-206/Domain: von Willebrand factor type A repeat homology <WMA1>  
 F127-262/Domain: EGF homology <EGF>  
 F127-437/Domain: von Willebrand factor type A repeat homology <WMA2>  
 F176-344/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F1221-228,234-247,249-262/Dissulfide bonds: #status predicted

Query Match 7.4%; Score 113.5; DB 2; Length 496;  
 Best Local Similarity 26.0%; Pred. No. 0.21;  
 Matches 59; Conservative 35; Mismatches 92; Indels 41; Gaps 13;

Qy 12 CYGG----FDVYILDKSGV-LHMNEIYFVEQLAHKFIISPOLRMSFTV--FSTR 61  
 Db 265 CSGGSGSSATDVLVFLIDGSKSVDPENFELVKFISQIVDTLVDKLAQVGLVQYSSSVR 324  
 Qy 62 GTTLMKLTEDREQRQGLEELQKLP--GGDTYMHGFERASEQIYENRQGRYTA 116  
 Db 325 QEPFGRRHTKDKIKAIVRN-----SYMERGTWTGALKYLINSFTVSSGARRPA 376  
 Qy 117 --VITATDGEHLDELFFYSREANRSDGAIYCVGVDPFNETOLARIADS--KDHVF 172  
 Db 377 QKQIVFTDGRSQD---YINDAKKAKDQFKRPFANGVGNVDELREINSEVAHYF 432  
 Qy 173 PVNDGFQALQGIHSITLKSCTEIIIAEPSTTCAGES--FOVVVRG 216  
 Db 433 VTAD-FRTINGIKKIQKICVE---BDP---CACESLVFQAKVEG 472

## RESULT 14

BBHU  
 Complement factor B precursor [validated] - human  
 N/Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-j  
 N/Contents: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment  
 C/Species: Homo sapiens (man)  
 C/Date: 19-Feb-1984 #sequence\_revision 05-Aug-1994 #text\_change 09-Jul-2004  
 C/Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628; I544  
 R/Melja, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.  
 submitted to the EMBL Data Library, March 1993  
 A/Reference number: S34075  
 A/Accession: S34075  
 A/Molecule type: mRNA  
 A/Residues: 1-764 <MEJ>  
 A/Cross-references: UNIPROT:P00751; EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g29756  
 R/Moore, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
 A/Title: Isolation of cDNA clones for the human complement protein factor B, a class III  
 A/Reference number: A44622; MUID:83039428; PMID:6957884  
 A/Accession: A44622  
 A/Molecule type: RNA  
 A/Residues: 467-546;550-595;752-764 <MOO>  
 A/Cross-references: GB:J00185; GB:J00186  
 A/Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation  
 R/Moie, J.B.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
 J. Biol. Chem. 259, 3407-3412, 1984  
 A/Title: Complete primary structure for the zymogen of human complement factor B.  
 A/Reference number: A20751; MUID:84161997; PMID:6546754  
 A/Accession: A00934  
 A/Molecule type: protein; mRNA  
 A/Residues: 26-764 <MOI>  
 A/Cross-references: GB:K01566  
 A/Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-V  
 A/Note: 736-Ser was also found  
 A/Note: glycosylation sites were determined  
 R/Christie, D.L.; Gagnon, J.  
 Biochem. J. 209, 61-70, 1983  
 A/Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of the

A:Reference number: A19188; MUID:83204002; PMID:6342610  
 A:Contents: the final paper in a series documenting the sequence, glycosylation site, an  
 A:Accession: A19188  
 A:Molecule type: protein  
 A:Residues: 260-296 'T', 298-764 <CHR>  
 R:Campbell, R.D.; Porter, R.R. 80, 4464-4468, 1983  
 Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
 A:Title: Molecular cloning and characterization of the gene coding for human complement  
 A:Reference number: A19947; MUID:83273641; PMID:6308626  
 A:Accession: A19947  
 A:Molecule type: DNA  
 A:Residues: 346-764 <C&M>  
 A:Cross-references: GB:U00125  
 A:Accession: B19947  
 A:Molecule type: mRNA  
 A:Residues: 339-509 <C&I>  
 A:Cross-references: GB:J00126; NID:9187723; PIDN:AAA56226.1; PID:9553536  
 R:Mu, L.; Morley, B.J.; Campbell, R.D.  
 Cell 48, 331-342, 1987  
 A:Title: Cell-specific expression of the human complement protein factor B gene: evidence  
 A:Reference number: A25971; MUID:87102880; PMID:3643061  
 A:Accession: B25971  
 A:Molecule type: DNA  
 A:Residues: 1-89 <MUL>  
 A:Cross-references: GB:M15082; NID:9187699; PIDN:AAA59625.1; PID:9553534  
 R:Niemann, M.A.; Brown, A.S.; Miller, E.J.  
 Biochem. J. 274, 473-480, 1991  
 A:Title: The principal site of glycation of human complement factor B.  
 A:Reference number: S14339; MUID:91174758; PMID:2006911  
 A:Accession: S14339  
 A:Molecule type: protein  
 A:Residues: 270-339 <NIB>  
 A:Note: binding site for carbohydrate to lysine under artificial conditions  
 R:Morley, B.J.; Campbell, R.D.  
 EMBO J. 3, 153-157, 1984  
 A:Title: Internal homologues of the B<sub>a</sub> fragment from human complement component factor B  
 A:Reference number: A4628; MUID:84158524; PMID:632161  
 A:Accession: A4628  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 16-225 'F', 227-259 <MOR>  
 R:Schwaeble, W.; Luttig, B.; Sokolowski, T.; Esteller, C.; Weiss, E.H.; Meyer zum Busche  
 Immunobiology 188, 221-232, 1993  
 A:Title: Human complement factor B: functional properties of a recombinant zymogen of th  
 A:Reference number: I54409; MUID:94041399; PMID:8225386  
 A:Accession: I54409  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-764 <RES>  
 A:Cross-references: GB:S67310; NID:9452937; PIDN:AD13989.1; PID:94261689  
 R:Horluch, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
 Mol. Immunol. 30, 1587-1592, 1993  
 A:Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conv  
 A:Reference number: I57824; MUID:94067177; PMID:8247029  
 A:Accession: I57824  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-31, 'O', 33-764 <RES>  
 A:Cross-references: GB:L15702; NID:9291921; PIDN:AAA16620.1; PID:9291922  
 C:Comment: 292-Cys has a free sulfhydryl.  
 C:Genetics:  
 A:Gene: GDB:BF  
 A:Cross-references: GDB:119726; OMIM:138470  
 A:Map position: 6p21.3-6p21.3  
 A:Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 65  
 A:Note: the list of introns may be incomplete  
 C:Complex: complement factor B initially forms an inactive complex with complement facto  
 ment factor C3b forming active C3/C5 convertase; B<sub>a</sub> is released  
 C:Function:  
 A:Description: B<sub>b</sub> is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha c  
 al  
 A:Pathway: complement alternate pathway

C:Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; v  
 C:Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydrol  
 F:1-25/Domain: signal sequence #status predicted <Sig>  
 F:26-764/Product: complement factor B #status experimental <MAT>  
 F:26-259/Product: complement factor B<sub>a</sub> fragment #status experimental <BAF>  
 F:37-98/Domain: complement factor H repeat homology <FH1>  
 F:103-158/Domain: complement factor H repeat homology <FH2>  
 F:165-218/Domain: complement factor H repeat homology <FH3>  
 F:260-764/Product: C3/C5 convertase B<sub>b</sub> fragment #status experimental <BBF>  
 F:268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
 F:482-752/Domain: trypsin homology #status atypical <TRY>  
 F:37-76-62-98-103-145-131-158-165-205-191-218-478-596-511-527-599-615-656-682-695-725/D  
 F:122,142,285,378/Binding site: carbohydrate (asn) (covalent) #status experimental  
 F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
 F:526-576,699/Active site: His, Asp, Ser #status experimental

Query Match 7.4%; Score 113.5; DB 1; Length 764;  
 Best Local Similarity 19.6%; Pred. No. 0.37;  
 Matches 66; Conservative 63; Mismatches 109; Indels 99; Gaps 17;

Qy 1 QGGRBDG-GPA-----CYGPFDFYILDKSGSV-----LHNMETYYFVQLAH 44  
 Db 243 EGVDAEDGHPGQKRTVLDPGSMNTYLVDSGSDSGASNFTGAKCLVNLIEKVAS 302  
 Qy 45 KPIFSPQIRMSFYFSTRGTTMLKLT---DRQIQGLLEL---QKVLPGSDTYMHG 96  
 Db 303 YGVKP--RYGLVYATYATPKIWMVSEADSSNADWVTKQLNEINYEDHKJSGTNT----- 355  
 Qy 97 FERASBGIVYENK-----QGY-RTASVYIALTDG-----ELHEDLFFPYSE 135  
 Db 356 -KQALQVIVSMGSPDVPPEGNNRTRVILMTDLHMGSPITVYIDIRDLVYIGKD 414  
 Qy 136 REANRSRDLCATYCVG--VKDFNETQLARADSKDHPVNDGFOALQGIHSILKSK 193  
 Db 415 RKNPREDYLDVYFVGPLVQVNVINALASKQNEQVFKVD--MENTLEDVFPQIMDES- 472  
 Qy 194 IEILAAPEPTICGSEFQVYVRNGFRHARNDVRLCSFKINDSVLNEKPSVEDYTL 253  
 Db 473 -----QSLSLC-----GMVWEHRGTD-----YHKPQWAKISV-- 501  
 Qy 254 CPAPILKEVGKMAALQVSMNDGLSFSSSVIITTC 290  
 Db 502 ----IRPSKHESCMG-----AVVSEYVLTALIC 527

RESULT 15  
 A54849  
 collagen alpha 1(VII) chain precursor - human  
 N:Alternate names: procollagen alpha 1(VII) chain  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Nov-1994 #sequence, revision 04-Nov-1994 #text, change 09-Jul-2004  
 A:Accession: A54849; PH0844; S16316; I56328; A20296; I84686  
 R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Utico, J.  
 J. Biol. Chem. 269, 20256-20262, 1994  
 A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha(VII)  
 A:Reference number: A54849; MUID:94327588; PMID:805117  
 A:Accession: A54849  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-2944 <CHR>  
 A:Cross-references: UNIPROT:Q02388; GB:L02870; NID:9987124; PIDN:AAA75438.1; PID:9987125  
 R:Tanaka, T.; Takahashi, K.; Futukawa, F.; Imanura, S.  
 Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
 A:Title: Molecular cloning and characterization of type VII collagen cDNA.  
 A:Reference number: PH0844; MUID:92231902; PMID:1567409  
 A:Accession: PH0844  
 A:Molecule type: mRNA  
 A:Residues: 1-340-475, 'RALSTASHSTLCWRATRHPCNRGSHWTRACBPCNRPAHRAAG', 524-528, 'C',  
 A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BA002853.1; PID:9453699  
 A:Experimental source: keratinocyte  
 A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
 R:Parente, M.G.; Chung, L.C.; Rymaszewski, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
 Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A>Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
 A/Reference number: S16316; MUID:91334580; PMID:1871109  
 A/Accession: S16316  
 A/Molecule type: mRNA  
 A/Residues: 815-892, 'E', 894-1439 <PAR>  
 A/Cross-references: GB:M65158; GB:649017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
 A/Experimental source: keratinocyte  
 R/Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Priyaanah, P.S.; Cook, M.E.; Wright, J.J.  
 J. Invest. Dermatol. 99, 691-696, 1992  
 A/Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot  
 A/Reference number: 156328; MUID:93107742; PMID:1469284  
 A/Accession: 156328  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 'ERR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
 A/Cross-references: GB:S51236; NID:G262308; PIDN:AA24637.1; PID:G262309  
 R/Beltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Ghanville, R.W.; Burgeson, R.E.  
 J. Biol. Chem. 264, 3822-3826, 1989  
 A/Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
 A/Reference number: A30296; MUID:89139437; PMID:2537292  
 A/Accession: A30296  
 A/Molecule type: Protein  
 A/Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'B', 2032, 'C', 2034-2041, '  
 A/Note: two reported peptides cannot be reliably located  
 R/Greenspan, D.S.  
 Hum. Mol. Genet. 2, 273-278, 1993  
 A/Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
 A/Reference number: 148103; MUID:93271985; PMID:8499916  
 A/Accession: 148103  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 2395-2871, 'S', 2873-2944 <RE2>  
 A/Cross-references: GB:L06862; NID:G388713; PIDN:AA89196.1; PID:G388714  
 R/Christiano, A.M.; Rymaszewski, M.; Uitto, J.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
 A/Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly--> Ser sub  
 A/Reference number: A55255; MUID:94224777; PMID:8170945  
 A/Accession: A55255  
 A/Comment: annotation  
 A/Content: Prolines and lysines at the third position of the tripeptide repeating unit (C  
 ed and subsequently O-glycosylated.  
 C/Genetics:  
 A/Gene: GDB:COL7A1; EBR1; EBD1; EB  
 A/Cross-references: GDB:128750; OMIM:120120  
 A/Map position: 3p21.3-3p21.3  
 A/Note: defects in this gene can result in dominant and recessive dystrophic epidermolys  
 A/Note: there are 118 introns  
 C/Complex: type VII collagen is probably a homotrimer  
 C/Function:  
 A/Description: structural component of extracellular polymer associated with anchoring f  
 C/Keywords: coll1; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli  
 F/1-16/Domain: signal sequence #status predicted <SIG>  
 F/17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
 F/17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
 F/36-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
 F/231-318/Domain: fibronectin type III repeat homology <FN1>  
 F/327-413/Domain: fibronectin type III repeat homology <FN2>  
 F/414-502/Domain: fibronectin type III repeat homology <FN3>  
 F/508-593/Domain: fibronectin type III repeat homology <FN4>  
 F/598-683/Domain: fibronectin type III repeat homology <FN5>  
 F/686-771/Domain: fibronectin type III repeat homology <FN6>  
 F/776-862/Domain: fibronectin type III repeat homology <FN7>  
 F/864-952/Domain: fibronectin type III repeat homology <FN8>  
 F/954-1045/Domain: fibronectin type III repeat homology <FN9>  
 F/1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F/1170-1172/Region: cell attachment (R-G-D) motif  
 F/1189-1253/Region: cysteine/proline-rich  
 F/1254-2783/Region: interrupted helical  
 F/1334-1336/Region: cell attachment (R-G-D) motif  
 F/2008-2010/Region: cell attachment (R-G-D) motif  
 F/2553-2555/Region: cell attachment (R-G-D) motif  
 F/2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
 F/2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <API>  
 F/337, 786, 1109/Binding site: carbohydrate (Aen) (covalent) #status predicted

F/2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper:  
 F/2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental  
 F/2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental  
 F/2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 7.38; Score 111.5; DB 2; Length 2944;  
 Best Local Similarity 24.11; Pred. No. 3.2;  
 Matches 52; Conservative 42; Mismatches 85; Indels 37; Gaps 11;  
 Qy 17 DLYEILDKSGSV-LHNMNEIYFPEOLAHKE---ISPO-LRMSPIVSTGTT---LMKL 68  
 Db 38 DIVFVLGGSSSIGNSNREVRSLFEGVLVLPFGAASQGVRFATVOYSDPPRFEGLDAL 97  
 Qy 69 TEDREQIRQGLIELQKVLPGSDTYMGEFPERASBOIYENROGYRTASVTIALTDGLME 128  
 Db 98 GSGGDVIR-AIRELS--YKGNRTGAALIHVDHVELPQLARPQGVKVCILITDGK-SQ 153  
 Qy 129 DLPFYSREANRSHDLCATYICGVQDFNETQLARLD--SKHVFPVNDGFQALQGIH 186  
 Db 154 DLV---DTAAQRUKGQGVKLFVAVGINKADPEELKRVASQPTSDFFFEVND-FSILRTILP 209  
 Qy 187 SILKSGI-----EILAEPS 203  
 Db 210 LVSRVCTTAGGVPTRPDDSTAPRDVLVSEPS 245

Search completed: June 13, 2005, 20:06:35  
 Job time : 21.0213 secs

**THIS PAGE BLANK (USPTO)**





RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Matanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RT Nucleic. Acids. Res. 32:40-45 (2004).  
RN [4]  
RN TISSUE: Kidney; (ISOFORM 4).  
RC MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RX Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shamen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Mazny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield V.S.N., Krzywinski M.I., Skalska J., Smalins D.E.,  
RA Scherch A., Schin J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [5]  
RN INTERACTION WITH ANTHRAX TOXIN.  
RP TISSUE: Placenta; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
RX MEDLINE=22608610; PubMed=14759258; DOI=10.1186/gb-2004-5-2-r8;  
RA Scobie H.M., Rathey G.J.A., Bradley K.A., Young J.A.T.,  
RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
RT receptor.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174 (2003).  
RL [6]  
RL SPLICING ISOFORM(S) THAT ARE POTENTIAL NMD TARGET(S).  
RN PubMed=14759258; DOI=10.1186/gb-2004-5-2-r8;  
RA Hillman R.T., Green R.E., Brenner S.B.,  
RT "An unappreciated role for RNA surveillance.";  
RT Genome Biol. 5:RESEARCH008.1-RESEARCH008.16 (2004).  
RL [7]  
RL FUNCTION: Cellular role is not yet known.  
CC [8]  
CC SUBUNIT: Binds to the protective antigen (PA) of *Bacillus*  
CC anthracis. Binding does not occur in the presence of calcium.  
CC [9]  
CC SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC [10]  
CC ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Experimental confirmation may be lacking for some  
CC isoforms;  
CC Name=1;  
CC IsoId=Q9H6X2-1; Sequence=Displayed;  
CC Note=May be produced at very low levels due to a premature stop  
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
CC Name=2;  
CC IsoId=Q9H6X2-2; Sequence=VSP\_000444, VSP\_000445;  
CC Name=3;  
CC IsoId=Q9H6X2-3; Sequence=VSP\_000446, VSP\_000447;  
CC Name=4;  
CC IsoId=Q9H6X2-4; Sequence=VSP\_000448, VSP\_000449;  
CC [11]  
CC TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells  
CC but not in normal endothelial cells.  
CC [12]  
CC DOMAIN: Binding to PA seems to be effected through the VWA domain.  
CC [13]  
CC SIMILARITY: Belongs to the ATR family.  
CC [14]  
CC SIMILARITY: Contains 1 VWA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL: AF729145; AKS2094.1; -;  
CC DR EMBL: AF421380; AAL26496.1; -;  
CC DR EMBL: AK025429; BAB15128.1; ALT\_INIT.  
CC DR EMBL: AK001463; BAA91707.1; ALT\_FRAME.  
CC DR EMBL: BC012074; AAI12074.1; -;  
CC DR Gene: HGNC:21014; ANTXR1.  
CC H-InvDB: HIX0002125; -;  
CC DR MIM: 606410; -;  
CC DR Interpro: IPR008400; Anth\_Ig.  
CC DR Interpro: IPR008399; Ant\_C.  
CC DR Interpro: IPR002035; VWF\_A.  
CC DR Pfam: PF05587; Anth\_Ig; 1.  
CC DR Pfam: PF05586; Ant\_C; 1.  
CC DR SMART: SM00327; VWA; 1.  
CC DR SMART: SM00327; VWA; 1.  
CC DR POSITIVE: PS50234; VWA; 1.  
CC KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
CC STGNL 1  
CC FT CHAIN 33  
CC FT DOMAIN 33  
CC FT TRANSSEM 322  
CC FT DOMAIN 343  
CC FT DOMAIN 44  
CC FT DOMAIN 360  
CC FT DOMAIN 506  
CC FT CARBOHYD 166  
CC FT CARBOHYD 184  
CC FT CARBOHYD 262  
CC FT VARSPPLIC 365  
CC FT VARSPPLIC 369  
CC FT VARSPPLIC 268  
CC FT VARSPPLIC 298  
CC FT VARSPPLIC 319  
CC FT VARSPPLIC 334  
CC FT SEQUENCE 564 AA; 62789 MW; B118A00D5DF2233 CRC64;  
CC Query Match 100.0%; Score 1526; DB 1; Length 564;  
CC Best Local Similarity 100.0%; Pred. No. 3.3e-112;  
CC Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 1 QGGRREDCGACVGGPDLVFTLTKSGSVLHNMETIYFVQVLAKHFTSPQLRMSTFYEST 60  
CC DB 28 QGGRREDCGACVGGPDLVFTLTKSGSVLHNMETIYFVQVLAKHFTSPQLRMSTFYEST 87  
CC QY 61 RGTTLMLTLDRCQIRGELQKVLPGCDTYNHEGFERASEQIYYENRGQYRTASYIA 120  
CC DB 88 RGTTLMLTLDRCQIRGELQKVLPGCDTYNHEGFERASEQIYYENRGQYRTASYIA 147  
CC QY 121 LTQDELHEDLFFYSERERANSRDIGALIVCVKDFENETOLARIADSKDHVPVNDGFOA 180  
CC DB 148 LTQDELHEDLFFYSERERANSRDIGALIVCVKDFENETOLARIADSKDHVPVNDGFOA 207  
CC QY 181 LOGTTHSLTKSCREIILAAEPSTTCAGESQVYVVRNGGPFHARVNRVLCSPKINDSVTL 240  
CC DB 208 LOGTTHSLTKSCREIILAAEPSTTCAGESQVYVVRNGGPFHARVNRVLCSPKINDSVTL 267  
CC QY 241 NEKFSVEDTYLLCPAPILKEVGKALKQVSNIDGLSFSSVITTTTHCSDG 293  
CC DB 268 NEKFSVEDTYLLCPAPILKEVGKALKQVSNIDGLSFSSVITTTTHCSDG 320

```
RESULT 2
A1R1_MOUSE STANDARD; PRT; 562 AA.
ID A1R1_MOUSE
AC Q9GC52;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Anthrax toxin receptor 1 precursor (tumor endothelial marker 8).
GN Name=Anthrax1; Synonyms=Atx1, Tem8;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NC1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21443268; PubMed=11559528;
RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,
RA Kinzler K.W., St Croix B.;
RT "Cell surface tumor endothelial markers are conserved in mice and
humans."
RL Cancer Res. 61:6649-6655 (2001).
RN [2]
RP SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Boro H., Kondo S.,
RA Nikiido I., Ogaso N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schirni L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Brad C., Brusic V., Chochia C., Corbani L.B., Cousins S.,
RA Dalla B., Dragan T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Giesel C., Godzik A., Gough J.,
RA Grifmond S., Guerlinch S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Karai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglocz D.R., Matras K., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numa K., Okido T., Pavan W.J., Petrea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Veratano R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmink L.G., Wyshaw-Borls A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kobayashi T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
CC -1- FUNCTION: Cellular role is not yet known.
CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus
anthracis (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=1;
CC IsoId=Q9GC52-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9GC52-2; Sequence=VSP_000450;
CC Note=No experimental confirmation available;
CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain
(by similarity).
CC -1- SIMILARITY: Belongs to the ATR family.
CC -1- SIMILARITY: Contains 1 VWA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AF378762; AAL1199.1; -.
CC EMBL; AK013005; BAB28591.1; ALT_INIT.
CC MGD; MGI:1916788; Anthx1.
CC InterPro; IPR008400; Anth_19.
CC InterPro; IPR008399; Ant_C.
CC InterPro; IPR02035; VWF_A.
CC Pfam; PF05587; Anth_1g; 1.
CC Pfam; PF05586; Ant_C; 1.
CC Pfam; PF00092; VWA; 1.
CC PROSITE; PS50234; VWF_A.
CC KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.
FT SIGNAL 1
FT CHAIN 30
FT DOMAIN 31 562
FT DOMAIN 31 319
FT TRANSMEM 320 340
FT DOMAIN 341 562
FT DOMAIN 42 213
FT DOMAIN 358 366
FT DOMAIN 501 562
FT CARBOHYD 164 164
FT CARBOHYD 182 182
FT CARBOHYD 260 260
FT VARSPIC 477 562
FT FT
FT FT
FT FT
SQ SEQUENCE 562 AA; 62308 MW; 6AC92043B4B4F7C CRC64;
Query Match 98.9%; Score 1509; DB 1; Length 562;
Best Local Similarity 99.0%; Pred. No. 7.3e-11;
Matches 289; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGRREDGPGACGCGFDLYFLIDKSGSVLHFMNRYVPEQLAKFTSPQRMSTIVSTR 61
DB 27 GGRREDGPGACGCGFDLYFLIDKSGSVLHFMNRYVPEQLAKFTSPQRMSTIVSTR 86
QY 62 GTTLMKLTREORQGLLELOKVLPGCDTYMHGEPFASQIYENRQGRTPASVITIAL 121
DB 87 GTTLMKLTREORQGLLELOKVLPGCDTYMHGEPFASQIYENRQGRTPASVITIAL 146
QY 122 TDGELHEDLFFYSBREANRSRDIGAIYVCVGDPMETQIARIADSKDHYFPVNDGFQAL 181
DB 147 TDGELHEDLFFYSBREANRSRDIGAIYVCVGDPMETQIARIADSKDHYFPVNDGFQAL 206
QY 182 QGIHSLTKSKCEIILAAEFTICAGSFGVVRVNGGFRARAVDRVLCSFKINDSVTLN 241
DB 207 QGIHSLTKSKCEIILAAEFTICAGSFGVVRVNGGFRARAVDRVLCSFKINDSVTLN 266
QY 242 EKPFSEVDYTLCPAPILKEVGKKAALQVSMNDGLSPFISSVITITTHCSGDG 293
DB 267 EKPFSEVDYTLCPAPILKEVGKKAALQVSMNDGLSPFISSVITITTHCSGDG 318
RESULT 3
A1R2_HUMAN STANDARD; PRT; 489 AA.
ID A1R2_HUMAN
AC P58335; O86U11; Q8NB13; Q96NC7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Anthrax toxin receptor 2 precursor (Capillary morphogenesis protein-2)
(GMG-2).
GN Name=ANTXR2; Synonyms=CMG2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
```

OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC MEDLINE=21539596; PubMed=11683410;  
 RX Bell S.E., Mavila A., Salazar R., Bayleee K.J., Kanagala S.,  
 RA Maxwell S.A., Davis G.E.;  
 RT "Differential gene expression during capillary morphogenesis in 3D  
 RT collagen matrices: regulated expression of genes involved in basement  
 RT membrane matrix assembly, cell cycle progression, cellular  
 RT differentiation and G-protein signaling.";  
 RL J. Cell Sci. 114:2755-2773(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH ANTHRAX TOXIN.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
 RA Scobie H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.;  
 RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
 RT receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
 RN [3]  
 RP SEQUENCE OF 78-489 FROM N.A. (ISOFORM 3), AND SEQUENCE FROM N.A.  
 RC (ISOFORM 4).  
 RX TISSUE=Synovial cell;  
 PubMed=14702039; DOI=10.1036/ngi1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Makametu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohyashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J.-i., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
 RA Nagahori A., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
 RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Puruya T., Kikawa E.,  
 RA Omura Y., Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M.,  
 RA Yamazaki M., Nishimura K., Ishibashi T., Yamashita H., Murakawa K.,  
 RA Fujimori K., Tanai H., Kimata S., Watanabe M., Hirooka S., Chiba Y.,  
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida S., Hoccu T.,  
 RA Kusano Y., Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O.,  
 RA Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K., Arita M.,  
 RA Imose N., Munesato K., Yuki H., Oshima A., Sasaki N., Aotaka S.,  
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki S., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujizawa T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Omori Y., Okamoto S.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Oktani R., Kawakami T., Noguchi N., Itoh T., Shigeta K., Senba T.,  
 RA Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togoashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Ilogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 CC -1- FUNCTION: Cellular role is not yet known.  
 CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 CC anthracis in a divalent cation-dependent manner, with the  
 CC following preference: calcium > manganese > magnesium > zinc.  
 CC Seems to bind to collagen type IV and laminin.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1 and 2).  
 CC Secreted (isoform 3). Isoform 1 is expressed at the cell surface  
 CC while isoform 2 is predominantly expressed within the endoplasmic  
 CC reticulum and not at the plasma membrane.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=P58335-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P58335-2; Sequence=VSP\_00843;  
 CC Name=3;  
 CC IsoId=P58335-3; Sequence=VSP\_008344, VSP\_008345;  
 CC Note=No experimental confirmation available;

CC	Name=4;	
CC	ISID=P58335-4; Sequence=VSP_008346;	
CC	Note=No experimental confirmation available;	
CC	-1- TISSUE SPECIFICITY: Expressed in colon, heart, kidney, lung,	
CC	liver, peritoneal blood leukocytes, placenta, skeletal muscle,	
CC	small intestine and spleen.	
CC	-1- DOMAIN: Binding to PA seems to be effected through the VWA domain.	
CC	-1- SIMILARITY: Belongs to the ATR family.	
CC	-1- SIMILARITY: Contains 1 VWA domain.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.fdb-sib.ch/announce/">http://www.fdb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).	
CC	-----	
DR	EMBL; AY040326; AK477222.1; -	
DR	EMBL; AY233452; AAP04016.1; -	
DR	EMBL; AK055636; BAB70976.1; ALT_INIT.	
DR	EMBL; AK091721; BAC03731.1; -	
DR	Genew; HGNC:21732; ANTXR2.	
DR	MTM; 608041; -	
DR	InterPro; IPR008399; Ant_C.	
DR	InterPro; IPR02035; VWF_A.	
DR	Pfam; PF00586; Ant_C.1.	
DR	Pfam; PF00092; VWA; 1.	
DR	PROSITE; PS50234; VWF_A; 1.	
KW	Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.	
FT	SIGNAL	1
FT	CHAIN	33
FT	DOMAIN	34
FT	DOMAIN	34
FT	TRANSMEM	319
FT	DOMAIN	342
FT	DOMAIN	44
FT	CARBOHYD	250
FT	CARBOHYD	260
FT	VARSPLIC	213
FT		315
FT	VARSPLIC	290
FT		322
FT		3)
FT		/FtId=VSP_008344.
FT		Missing (in isoform 3).
FT	VARSPLIC	323
FT		489
FT	VARSPLIC	477
FT		489
FT		4)
FT		/FtId=VSP_008346.
FT	CONFLICT	357
FT		357
FT	SEQUENCE	489 AA; 53692 MW; B9F679DB5B6E2B7 CRC64;
QY	Query Match	51.5%; Score 786; DB 1; Length 489;
QY	Best Local Similarity	53.9%; Pred. No. 9.6e-54;
QY	Matches 153; Conservative	54; Mismatches 75; Indels 2; Gaps 1;
DB	10 PACVGGDPLFDLDDSGSVLHNNHIEYFYFQALHKEFISPOLRMSFIVSFRGTLMKLT	69
DB	37 PSRCRAADPLFDLDDSGSVANMTEINFPQALERKRVSPEMRSLFVSSQATIIILPLT	96
QY	70 EDRBOIQGLLELOKVLPGGDYVNHGSEFRASEQIYYENOGYRTASVIAIITDELHED	129
DB	97 GDRGKISKGLIEDLKRVPVGGTYIHGKLKLANEQI--QKAGGLKTSIIILITLGGKLDGL	154
QY	130 LFFPSEENARSRGLAIYVCVGRKDPNENPOLAIASKDHVPFVNDGFOALQIIHSIL	189
DB	155 VPSVAEERAKISRLGASVYCVGLADEQAQLERIASKQVFPVKGQFQALKGIINSIL	214
QY	190 KKSCEIILAAEPTTCAGSEFOVVNRANGRHHANVDRVICSFKINISVTLNEXRPSVED	249
DB	215 AQSCTELIELQPSVVCVGBERQIYLSRGRFGLSRNSVSLCTYVNNETVTSVAPVQVL	274
QY	250 TYLLCPAPILKEVGMKALQVSNMDGLSFLISSVITITTHCSOG	293

Db 275 NSMLCPAPILNKAGETLDOVSFNGGKSVIGSLIVTATECSNG 318

RESULT 4

06DPX2 PRELIMINARY; PRT; 487 AA.

AC 06DPX2; 25-OCT-2004 (TREMBlrel. 28, Created)

DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)

DE Anthrax toxin receptor 2.

GN Name=Anthr2;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

[1]

SEQUENCE FROM N.A.

RA STRAIN=C57BL/6; TISSUE=Eye;

RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datchenko L., Marubina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Uddin T.B., Toshlyuk S., Carninci P., Prange C.,

RA Baha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richard S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huliyk S.W.,

RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,

RA Paley J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalske U., Smallegre D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Maira M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Eye;

RA Strauberg R.;

RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC076595; AAT76595.1.;

DR GO: GO:0016021; C:Integral to membrane, IEA.

DR GO: GO:0004872; F:receptor activity, IEA.

DR InterPro: IPR008400; Anth\_IG.

DR InterPro: IPR008399; Ant\_C.

DR InterPro: IPR002035; VWF\_A.

DR Pfam: PF05587; Anth\_IG; 1.

DR Pfam: PF05586; Ant\_C; 1.

DR Pfam: PF00092; VWA; 1.

DR SMART: SM00337; VWA; 1.

DR PROSITE: PS50234; VWFA; 1.

KM Receptor.

SO SEQUENCE 487 AA; 53184 MW; 61A400D60BCDE69 CRC64;

Query Match 50.5%; Score 770; DB 2; Length 487;

Best Local Similarity 52.8%; Pred. No. 1.8e-52;

Matches 150; Conservative 52; Mismatches 80; Indels 2; Gaps 1;

Qy 10 PACGSGDYLPIIDKSSGLAHNNEIYFPEQLAKPISPOLRNSFIYFSGRTTLMKLT 69

Db 37 PSCKKADPLVFLDKSSGVANWIEIYFVQLTERFVSPMRISFIVFSSQATIIILPLT 96

Qy 70 EDREQIQGLLELDKIVLPGSDTYVHGEGFERASQEIYYENRGYRTASVITALTGGLHED 129

Db 97 GDRKIKIGKLELDKAVKPVGSETTYIHGKILANEQI--QNAQGLKASSIIILTGKLDGL 154

Qy 130 LFYSEREANRSRDGLAIYVCVGVKDFNETQLARIADSKOHVFPVNDGFOALQGIHSIL 189

Db 155 VPSYAENBAKKSRSIGASVVCVGLDFEQALERIADSKQVFPVKGCFQALKGIHSIL 214

Qy 190 KKSCEITLAAEPSTICGESFOVYVNGNPFHANVDRVYCSFKINSVLTNKEPFSVED 249

Db 215 AQSCETILELSPSSVCGEKFQVLTGRAVTSISHDSGLTPTANSTYTKSRPVSTQP 274

Qy 250 TYLLCPAPILKENVGKKAALQVSNMDGLSPISSVYITTTTCSDG 293

Db 275 SSILCPAPVINKOETILEVTSISYNDGKANSRSRLITRATCTG 318

RESULT 5

08BVM2 PRELIMINARY; PRT; 641 AA.

AC 08BVM2; 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Mus musculus adult male testis cDNA, RIKEN full-length enriched

DE library, clone:493430v1 product:hypothetical Proline-rich region/von

DE Willebrand factor type A domain containing protein, full insert

DE sequence.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

[1]

SEQUENCE FROM N.A.

RA STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.;

RT "High-efficiency full-length cDNA cloning."

RL Meth. Enzymol. 303:19-44(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RA The FANTOM Consortium.

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Komno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalization and subtraction of cap-trapper-selected cDNAs to

RT prepare full-length cDNA libraries for rapid discovery of new genes."

RL Genome Res. 10:1617-1630(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Testis;

RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,

RA Komno H., Akiyama Y., Nishi K., Kitanai T., Tashiro H., Itoh M.,

RA Smit N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

RA Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,

RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

RT "RIKEN integrated sequence analysis (RISA) system-384-format

RT sequencing pipeline with 384 multicapillary sequencer."

RL Genome Res. 10:1757-1771(2000).

RN [6]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA Adachi J., Aizawa K., Akimura T., Arikawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashimoto W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiroka T., Hirose T.,  
 RA Hori P., Imocani K., Ishii Y., Itoh M., Kagawa I., Kanakawa T.,  
 RA Katoh H., Kawai Y., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Saeki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takano-Akiba S., Tanaka T., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK077206; BAC36683.1; -  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR008400; Anth\_Ig.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF05587; Anth\_Ig; 1.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS50234; VWF\_A; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 641 AA; 70415 MW; 199E300730BC85E3 CRC64;  
 Query Match 39.7%; Score 606; DB 2; Length 641;  
 Best Local Similarity 42.2%; Pred. No. 2,3e-39;  
 Matches 119; Conservative 58; Mismatches 105; Indels 0; Gaps 0;  
 QY 9 GPACGAGFDLYPLIDKSGSVLHNNELIYFVBOQLAHKFIISQLRNFIIVSTRTTLMKL 68  
 DB 68 GDCCGIGFDLYPLIDKSGSVADNWHIYFAGLVKCKFTNNLRISITVSTEARVILPL 127  
 QY 69 TEDREIQGLJELQKVLPGDVTYHGEFERASBOIYENNGQYRTASVIALDGEIHE 128  
 DB 128 TSDSEIKNSLLVLSKIVPGGLTHMQKGLRKANEGIRSTLGGRIIVSVIALLDGLLL 187  
 QY 129 DLFFYSERANRSRLGAIYVCVGVDFNETOLARIADSKDHFVPVNDFOALOGIHSI 188  
 DB 188 KPYLDITMEAKKARMGAIYTVGVPMYSKQOLVNIADPRRCVDDGFGALGVNDPL 247  
 QY 189 LKKSCEILAAEPSTICAGESFOVVVRNGFRHARNVRLVCSFKINDSVTLNKPESVE 248  
 DB 248 TSKSCTEILSVQFTYVCAKDFYQVNIISGHLNNTSMKQVLCRFKSDSKVVDSPDMN 307  
 QY 249 DTYLLCPAPILKEVGMKALQVSMNDGSLFSSSVIITTHC 290  
 DB 308 EHSITCPGPKIKHGEDVSLQVSLNNGISFLGNKLITSTNC 349

RESULT 6  
 Q96EC6 PRELIMINARY; PRT; 97 AA.  
 AC Q96EC6;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE ANTXR1 protein (Fragment).  
 GN Name=ANTXR1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast;  
 RX MEDLINE=2338257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheef C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cassavet T.L., Schetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toehlyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Skalko U., Smalins D.E., Schnerch A., Schein J.E.,  
 RA Krzyzanski M.I., Skalka U., Smalins D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast;  
 RA Strausberg R.;  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC012475; AAH12475.1; -  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro; IPR008400; Anth\_Ig.  
 DR Pfam; PF05587; Anth\_Ig; 1.  
 DR NON TER 1  
 SQ SEQUENCE 97 AA; 10453 MW; 14F475F0B170E71A CRC64;  
 Query Match 25.4%; Score 388; DB 2; Length 97;  
 Best Local Similarity 95.1%; Pred. No. 4.1e-22;  
 Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 210 FOVVVRNGFRHARNVRLVCSFKINDSVTLNKPESVEDTYLLCPAPILKEVGMKALQ 269  
 DB 1 FOVVVRNGFRHARNVRLVCSFKINDSVTLNKPESVEDTYLLCPAPILKEVGMKALQ 60  
 QY 270 VSMNDGSLFSSSVIITTHCS 291  
 DB 61 VSMNDGSLFSSSVIITTHCS 82

RESULT 7  
 ITAD RAT STANDARD; PRT; 1161 AA.  
 ID ITAD RAT  
 AC Q99YB7;  
 DT 05-JUL-2004 (Rel. 44, Created)  
 DT 05-JUL-2004 (Rel. 44, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Integrin alpha-D precursor.  
 GN Name=Itgad;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA O'Brien M.M., VanderVlieten M., Klugman P.D., Dietrich G.,  
 RA Gallatin W.M.;  
 RT "Cloning of rat alpha D, a novel beta 2 integrin.";  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1 FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and  
 CC VCAM1. May play a role in the atherosclerotic process such as  
 CC clearing lipoproteins from plaques and in phagocytosis of blood-  
 CC borne pathogens, particulate matter, and senescent erythrocytes  
 CC from the blood (By similarity).  
 CC -1 SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D  
 CC associates with beta-2 (By similarity).  
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
 CC -1 DOMAIN: The integrin I-domain (inset) is a VWF domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -1 SIMILARITY: Belongs to the integrin alpha chain family.

```

CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VMPA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.emb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF021334; AAF21241.1; -.
DR HSPB; P1215; 1BRQ.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR02035; VMP_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00453; VMPADOMAIN.
DR SMART; SM00337; VMA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VMPA; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
DR Repeat; Signal; Transmembrane.
KW SIGNAL.
FT CHAIN. 1 19 Potential.
FT TRANSMEM 20 1161 Integrin_alpha-D.
FT DOMAIN 1101 1121 Extracellular (Potential).
FT DOMAIN 1122 1161 Potential.
FT REPEAT 34 87 Cytoplasmic (Potential).
FT REPEAT 88 87 FG-GAP 1.
FT REPEAT 152 334 VMPA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.
FT CA_BIND 467 475 Potential.
FT CA_BIND 531 539 Potential.
FT CA_BIND 594 602 Potential.
FT SITE 1126 1130 GFPKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 712 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;
Query March 10.5%; Score 159.5; DB 1; Length 1161;
Best Local Similarity 27.9%; Pred. NO. 0.00098;
Matches 63; Conservative 41; Mismatches 93; Indels 29; Gaps 11;
Qy 17 DLYEILDKSGSV-LHHNWEIYFVEQLAKRISPOLRMSFVFPSTRTGLTKLTE----- 70
Db 152 DIATLIQSSGSSINGRDFAQMDFVYKALMGSRFASSTLFLSMQYINILKTHFTFEPKNIL 211
Qy 71 DREQIRQGLEBLQKLPQGGDTYMEGFERASQIYENRQGYRTA-SVIALLTQGLHEHD 129
Db 212 DQSLVDPPIVQLQ-----GLTYTATGIRTVMEBELFHSKNGSKSAKKILVITDQOKYRD 266

```

```

Qy 130 LFPYSE--REANRSRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPVNDGROA 180
Db 267 PLKSDVLPADAKA---GIIRYALGVGDAPQEPALKEMLNTIGAPQDHFYKGN-FAA 322
Qy 181 LQGIHSHILKKSCEIILAEPSITCAGESFQVYVNGGFRRARVD 226
Db 323 LRSIQRLQEK-----IFALBCTQSRSSSFQHEMSQGFSSALTSND 364
RESULT 8
ITAD HUMAN STANDARD; PRT; 1162 AA.
AC Q13349; Q15575; Q15576;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor (LeukoIntegrin alpha D) (CD11d) (ADB2).
GN Name=ITGAD;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=96111956; PubMed=8777714; DOI=10.1016/1074-7613(95)90058-6;
RA Van der Vlieten M., Le Trong H., Wood C.L., Moore P.F., St John T.,
RT Staunton D.E., Gallatin W.M.;
RL "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-3."
RN [2]
RP IMMUNITY 3:683-690(1995).
RX SEQUENCE OF 1-235 FROM N.A.
RX MEDLINE=20187620; PubMed=10722744; DOI=10.1074/jbc.275.12.8959;
RA Noci J.D., Johnson A.K., Dillon J.D.;
RT "Structural and functional characterization of the leukocyte integrin gene CD11d. Essential role of Spi and Sp3."
RL J. Biol. Chem. 275:8959-8969(2000).
RN [3]
RP SEQUENCE OF 467-537, 571-602, 633-666, 788-834 AND 910-1125 FROM N.A.
RX MEDLINE=96257236; PubMed=8666289; DOI=10.1016/0378-1119(95)00869-1;
RA Wong D.A., Davis E.M., Lebeau M., Springer T.A.;
RT "Cloning and chromosomal localization of a novel gene-encoding a human beta 2-integrin alpha subunit."
RL Gene 171:291-294(1996).
RN [4]
RP INTERACTION WITH VCAM1.
RX MEDLINE=99059842; PubMed=9841932;
RA Grayson M.H., Van der Vlieten M., Sterbinsky S.A., Michael Gallatin W.,
RT Hoffman P.A., Staunton D.E., Bochner B.S.;
RL "alpha2beta2 integrin is expressed on human eosinophils and functions as an alternative ligand for vascular cell adhesion molecule 1 (VCAM-1)."
RN [5]
RP J. Exp. Med. 188:2187-2191(1998).
RX INTERACTION WITH VCAM1.
RX MEDLINE=99370002; PubMed=10438935;
RA Van der Vlieten M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,
RT Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;
RL "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a binding interface between I domain and VCAM-1."
RN J. Immunol. 163:1984-1990(1999).
RX FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and VCAM1. May play a role in the atherosclerotic process such as clearing lipoproteins from plaques and in phagocytosis of blood-borne pathogens, particulate matter, and senescent erythrocytes from the blood.
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D associates with beta-2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed moderately on myelomonocytic cell lines and subsets of peripheral blood leukocytes and strongly on

```

CC tissue-specialized cells, including macrophages foam cells within  
 CC atherosclerotic plaques, and on splenic red pulp macrophages.  
 CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 CC EMBL: U37028; AAB36547.1; -;  
 CC EMBL: U40274; AAB60634.1; -;  
 CC EMBL: U40275; AAB60635.1; -;  
 CC EMBL: U40276; AAB60636.1; -;  
 CC EMBL: U40277; AAB60637.1; -;  
 CC EMBL: U40279; AAB60638.1; -;  
 CC EMBL: U40278; AAB60638.1; JOINED.  
 CC EMBL: AF187881; AAF62875.1; -;  
 CC HSSP: P11215; 1BHQ.  
 CC Genew: HGNC:6146; ITGAD.  
 CC MIM: 602453; -;  
 CC GO: GO:0008305; C:integrin complex; TAS.  
 CC GO: GO:0016337; P:cell-cell adhesion; NAS.  
 CC GO: GO:0007160; P:cell-matrix adhesion; NAS.  
 CC GO: GO:0006955; P:immune response; NAS.  
 CC InterPro: IPR000413; Integrin\_alpha.  
 CC InterPro: IPR02035; VWF\_A.  
 CC Pfam: PF01839; FG-GAP; 3\_  
 CC Pfam: PF00357; Integrin\_alpha; 1.  
 CC Pfam: PF00092; VWA; 1.  
 CC PRINTS: PRO1185; INTEGRINA.  
 CC PRINTS: PRO0453; VWFADOMAIN.  
 CC SMART: SM00191; Int\_alpha; 5.  
 CC SMART: SM00327; VMA; 1.  
 CC PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 CC PROSITE: PS50234; VWFA; 1.  
 CC KEGG: Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
 CC Repeat; Signal; Transmembrane.  
 CC CHAIN: 1 17 Potential.  
 CC DOMAIN: 18 1162 Extracellular (Potential).  
 CC TRANSMEM: 1101 1121 Potential.  
 CC DOMAIN: 1122 1162 Cytoplasmic (Potential).  
 CC REPEAT: 32 85 FG-GAP 1.  
 CC REPEAT: 86 ? FG-GAP 2.  
 CC DOMAIN: 150 332 VWFA.  
 CC REPEAT: 350 400 FG-GAP 3.  
 CC REPEAT: 401 452 FG-GAP 4.  
 CC REPEAT: 454 516 FG-GAP 5.  
 CC REPEAT: 518 576 FG-GAP 6.  
 CC REPEAT: 581 633 FG-GAP 7.  
 CC REPEAT: 465 473 Potential.  
 CC CA\_BIND: 530 538 Potential.  
 CC CA\_BIND: 593 601 Potential.  
 CC SITE: 1127 1131 GPFKR motif.  
 CC DISULFID: 67 74 By similarity.  
 CC DISULFID: 106 124 By similarity.  
 CC DISULFID: 655 710 By similarity.  
 CC DISULFID: 769 775 By similarity.  
 CC DISULFID: 846 861 By similarity.  
 CC DISULFID: 994 1018 By similarity.  
 CC DISULFID: 1023 1028 By similarity.  
 CC CARBOHYD: 59 59 N-linked (GlcNAc... ) (Potential).  
 CC CARBOHYD: 87 87 N-linked (GlcNAc... ) (Potential).  
 CC CARBOHYD: 99 99 N-linked (GlcNAc... ) (Potential).  
 CC CARBOHYD: 391 391 N-linked (GlcNAc... ) (Potential).  
 CC CARBOHYD: 691 691 N-linked (GlcNAc... ) (Potential).

FT CARBOHYD 733 733 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 873 873 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 957 957 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1046 1046 N-linked (GlcNAc... ) (Potential).  
 FT CONFLICT 500 500 Missing (in Ref. 2).  
 FT CONFLICT 515 518 GHW -> ATP (in Ref. 2).  
 FT CONFLICT 825 825 L -> V (in Ref. 2).  
 FT CONFLICT 984 984 V -> A (in Ref. 2).  
 SQ SEQUENCE 1162 AA; 12685 MW; F296A1A35455D77D CRC64;  
 Query Match 10.2%; Score 156; DB 1; Length 1162;  
 Best Local Similarity 24.3%; Pred. No. 0.0019;  
 Matches 67; Conservative 50; Mismatches 119; Indels 40; Gaps 11;  
 QY 7 DGGPAC-YGGFDLYFLIDKSGSV-LHRMWEIYFVQLAHKFTSPQLRMSFVSTRTGT 64  
 DB 139 DATPECPHQEMDVLVFLDGSIDQNDPMKGFVQAVMQFGDTLFLALMQYS-----N 194  
 QY 65 LMKLTEDREQIRGLBELQVLP-----GGDTYHBEFERSBQIYYENRGYRTA-SVIT 119  
 DB 195 LKIHFTFTQFRTPSQQSIVDPIVQLKGLFTPATGILTVVTLFHHKNGARRSAKKILI 254  
 QY 120 ALTDGELHEDLPYSSREANRSRLGAIYVCVGDFFNETQLAR-----IDSKDHF 172  
 DB 255 VITDQKXKDPLEYSD-VLPQAEKAGIRYAVGVGAFCQPTKROELNTISSAPPDHF 313  
 QY 173 PVNDGFQALGIIHSLIKSCIEILAEPSITCAGESFQVVRNGFRHARNVDVLC-- 230  
 DB 314 KV-DNFALLGSIKQIQEK-----IYAVEGTQSASSSFQHEMQEGHFTALMDGLFLGA 368  
 QY 231 --SPKINDS-----VTLNEKPSVEPTYL 252  
 DB 369 VGSFWSWGAFLYPNNMSPFTINMQENVDMDRSTYL 404  
 RESULT 9  
 Q9BP08 PRELIMINARY; PRT; 1332 AA.  
 ID Q9BP08  
 AC Q9BP08  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Integrin alpha H1 precursor.  
 GN Name=HITGAL;  
 OS Halocynthia roretzi (Sea squirt).  
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
 OC Stolidobranchia; Pyridae; Halocynthia.  
 OX NCBI\_TaxID=7729;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hemocyte;  
 RX MEDLINE=21103187; PubMed=11160215;  
 RA Miyazawa S., Azumi K., Nonaka M.;  
 RT "Cloning and characterization of integrin alpha subunits from the  
 RT solitary ascidian, Halocynthia roretzi.";  
 RL J. Immunol. 166:1710-1715(2001).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC EMBL: AB048261; BAB21479.1; -;  
 CC HSSP: P11215; 1BHQ.  
 CC GO: GO:0016021; C:integral to membrane; IEA.  
 CC GO: GO:0008305; C:integrin complex; IEA.  
 CC GO: GO:0005515; F:protein binding; IEA.  
 CC GO: GO:0007160; P:cell-matrix adhesion; IEA.  
 CC GO: GO:0007229; P:integrin-mediated signaling pathway; IEA.  
 CC InterPro: IPR000413; Integrin\_alpha.  
 CC InterPro: IPR02035; VWF\_A.  
 CC Pfam: PF00092; VMA; 1.  
 CC PRINTS: PRO1185; INTEGRINA.  
 CC PRINTS: PRO0453; VWFADOMAIN.  
 CC SMART: SM00191; Int\_alpha; 5.  
 CC SMART: SM00327; VMA; 1.  
 CC PROSITE: PS50234; VWFA; 1.



KW Cell adhesion; Integrin; Signal; Transmembrane.  
FT SIGNAL 1 30 Potential.  
FT CHAIN 31 1332 Integrin alpha H1.  
SQ SEQUENCE 1332 AA; 145851 MW; 0D9108D2B05CFEAE CRC64;  
Query Match 10.1%; Score 153.5; DB 2; Length 1332;  
Best Local Similarity 24.1%; Pred. No. 0.0035;  
Matches 63; Conservative 42; Mismatches 81; Indels 75; Gaps 10;  
QY 2 GGRREDG-----PACYGFDLYFLIDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFI 56  
DB 188 GNMREGSRNTECPG--SGVDLFLVLDGSGSVGRKFDKVKMVKKIT----- 232  
QY 57 VFSTRGTTLMKLTEDREQLRGLBELQKVLPG-----GDTYMHGEPERAS 101  
DB 223 -----AKLDIKEIVRGVGVQSHYVEGKSINKQYITTEISIGFKLIDFENAV 283  
QY 102 EQLIYENRGYRT-----ASVIALTDGELHEDLFFYSREARNS 141  
DB 284 DRI--QLQGYTTYGRALQKVIKRDPAIYGNKQVLLLTLDGQAKDKILP--NARL 338  
QY 142 RDLGAIYVCVGVDFNETOLARIA--DSKDHVPVNDGFQALQGIHSLKSCIEILA 198  
DB 339 RAKGATRAVGVGEIDISELKLIASGTSTDRVFTVTD--FGELDSIVSLQTEIQSFVLE 397  
QY 199 AEPSTICAGESFQVVRNGNP 219  
DB 398 CGKSAKTAG--YEMHFGENG 416  
RESULT 10  
Q8NFW1  
ID Q8NFW1 PRELIMINARY; PRT; 1626 AA.  
AC Q8NFW1;  
DT 01-OCT-2002 (TREMblrel. 22, Created)  
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE Alpha 1 type XXII collagen.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cartilage;  
RA Koch M., Jin M., Ashworth T., Burgess R.E.;  
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF406780; AA003620.1; -.  
DR HSBP; F18614; IIMP.  
DR GeneW; HGNC:22989; COL22A1.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0007155; P:cell adhesion; IEA.  
DR GO; GO:0006817; P:phosphate transport; IEA.  
DR InterPro; IPR008161; C1g\_helix.  
DR InterPro; IPR008160; C1g\_helix.  
DR InterPro; IPR008985; Consilike\_1ec\_g1.  
DR InterPro; IPR003129; TSP\_N.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01391; Collagen\_16.  
DR Pfam; PF00092; VMA; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR PRODOM; PD000007; C1g\_helix; 7.  
DR SMART; SM00210; TSPN\_1.  
DR SMART; SM00327; VMA; 1.  
DR PROSITE; PS50234; VMA; 1.  
KW Collagen.  
SQ SEQUENCE 1626 AA; 161115 MW; 34C68E3CFD467407 CRC64;  
Query Match 9.7%; Score 148.5; DB 2; Length 1626;  
Best Local Similarity 25.4%; Pred. No. 0.011;  
Matches 65; Conservative 45; Mismatches 105; Indels 41; Gaps 14;

QY 8 GGPACYG-----GFDLYFLIDKSGSV-LHHMNEIYFVEQLAHKF-ISP-QLRMSFI 56  
DB 21 GGGGCGAQRACCKSVHYDLVFLDTSSSVGKEDEKVKQVAVNVDTREVGPDRTRGVV 80  
QY 57 VFSTRGTTLMK--LTEDREQLRGLBELQKVLPGDPTYMHGEPERASQIYENRG--- 111  
DB 81 RYSDPPTTAFELGPGSQEEVKAARL--AYHGANTTGDALRYTARSFSPHAGGPR 138  
QY 112 ---YRTASVIALTDGELHEDLFFYSREARNSRDLGAIYVCVGVDFNETOLARIAD-- 166  
DB 139 DRAK--QVALLTDGR--SGLVLDAAAAARA---GIRPAGVGLAKLELEISAP 192  
QY 167 SKDHVPVNDGFQALQGIHSLKSCIEILAEPSTICAGESFQVVRNGNPFRRARND 226  
DB 193 KSAHVFHVSD--FNALDKIRGLRRLCENVLC--PS-----VRVEGDRFKHTNGT 240  
QY 227 RVLCSPKINDSVTINE 242  
DB 241 KEITGFDLMDLFSYKE 256  
RESULT 11  
Q8T6U5  
ID Q8T6U5 PRELIMINARY; PRT; 441 AA.  
AC Q8T6U5;  
DT 01-JUN-2002 (TREMblrel. 21, Created)  
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMblrel. 24, Last annotation update)  
DE Proximal thread matrix protein 1 variant a.  
OS Mytilus edulis (Blue mussel).  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
OC Mytiloidea; Mytilidae; Mytilus.  
OX NCBI\_TaxId=6550;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;  
RA Sun C., Lucac J.M., Waite J.H.;  
RT "Collagen-binding matrix proteins from elasmobranch extraorganic  
RT bursal fibers."  
RL Biomacromolecules 3:1240-1248(2002).  
DR EMBL; AF414454; AA083537.1; -.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00092; VMA; 2.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VMA; 2.  
DR PROSITE; PS50234; VMA; 2.  
KW Matrix protein.  
SQ SEQUENCE 441 AA; 47543 MW; 881D8BD36B891D2B CRC64;  
Query Match 9.7%; Score 148; DB 2; Length 441;  
Best Local Similarity 25.5%; Pred. No. 0.0025;  
Matches 51; Conservative 46; Mismatches 77; Indels 26; Gaps 12;  
QY 12 CYGFDLYFLIDKSGSV-----LHHMNEIYFVEQLAHKF--ISP-LRMSFIVSTRGTT 64  
DB 236 CAGHADIAPFDVASSINANNPNYGLMKDFMCDIVRFKGTGPDGQFAVVTADADATK 295  
QY 65 LMKLTE--DREQLRGLBELQKVLPG--GDTYMHGEPERASBOIYENRG---YRTASV 117  
DB 296 QFGIKDYSSKAEIKGAID---KTPSIIGQTALIDGLENARLEV--FNNRGGGREGVQKV 351  
QY 118 IIALTDGEL--HEDLFFYSREARNSRDLGAIYVCVGV--KDFNETOLARIADSDHYFPV 174  
DB 352 VILLTDQNNNGHS---PEHESLNRKEGVVAIVGVGTGLKSELINIASSEIVF--T 406  
QY 175 NDGFQALQGIHSLKSCI 194  
DB 407 TSSFNKLKIMENVVYKACM 426  
RESULT 12  
Q8T5C3



RT assessment." ;  
 RL Structure 6:923-935(1998).  
 RN [12]  
 RP 3D-STRUCTURE MODELING OF 17-616.  
 RX MEDLINE=98226734; PubMed=9560195; DOI=10.1073/pnae.95.9.4870;  
 RA Oxyg C., Springer T.A.;  
 RT "Experimental support for a beta-propeller domain in integrin alpha-subunits and a calcium binding site on its lower surface." ;  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
 CC -1- FUNCTION: Integrin alpha-M/beta-2 is implicated in various adhesive interactions of monocytes, macrophages and granulocytes as well as in mediating the uptake of complement-coated particles. It is identical with CR-3, the receptor for the IC3b fragment of the third complement component. It probably recognizes the R-G-D peptide in C3b. Integrin alpha-M/beta-2 is also a receptor for fibrinogen, factor X and ICAM1. It recognizes P1 and P2 peptides of fibrinogen gamma chain.  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-M associates with beta-2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and granulocytes.  
 CC -1- DOMAIN: The integrin I-domain (insert) is a WFPA domain. Integrins with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 WFPA domain.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD11b entry;  
 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/announcements or send an email to license@ebi.ac.uk).  
 CC -----  
 DR EMBL J03925; AA059544.1; -;  
 DR EMBL M18044; AA059491.1; -;  
 DR EMBL J04145; AA059903.1; -;  
 DR EMBL S52227; AAB24821.1; -;  
 DR EMBL S52152; AAB24821.1; JOINED.  
 DR EMBL S52153; AAB24821.1; JOINED.  
 DR EMBL S52154; AAB24821.1; JOINED.  
 DR EMBL S52155; AAB24821.1; JOINED.  
 DR EMBL S52157; AAB24821.1; JOINED.  
 DR EMBL S52159; AAB24821.1; JOINED.  
 DR EMBL S52161; AAB24821.1; JOINED.  
 DR EMBL S52164; AAB24821.1; JOINED.  
 DR EMBL S52165; AAB24821.1; JOINED.  
 DR EMBL S52167; AAB24821.1; JOINED.  
 DR EMBL S52169; AAB24821.1; JOINED.  
 DR EMBL S52170; AAB24821.1; JOINED.  
 DR EMBL S52173; AAB24821.1; JOINED.  
 DR EMBL S52174; AAB24821.1; JOINED.  
 DR EMBL S52180; AAB24821.1; JOINED.  
 DR EMBL S52181; AAB24821.1; JOINED.  
 DR EMBL S52184; AAB24821.1; JOINED.  
 DR EMBL S52189; AAB24821.1; JOINED.  
 DR EMBL S52191; AAB24821.1; JOINED.  
 DR EMBL S52192; AAB24821.1; JOINED.  
 DR EMBL S52203; AAB24821.1; JOINED.  
 DR EMBL S52212; AAB24821.1; JOINED.  
 DR EMBL S52213; AAB24821.1; JOINED.  
 DR EMBL S52216; AAB24821.1; JOINED.  
 DR EMBL S52219; AAB24821.1; JOINED.  
 DR EMBL S52220; AAB24821.1; JOINED.  
 DR EMBL S52221; AAB24821.1; JOINED.  
 DR EMBL S52222; AAB24821.1; JOINED.  
 DR EMBL S52226; AAB24821.1; JOINED.  
 DR EMBL M76724; AA058410.1; -;  
 DR EMBL M84477; AA051960.1; -;

DR PIR; A31108; RWHUB.  
 DR PDB; 1ABX; Model; @=17-1152.  
 DR PDB; 1BHO; X-ray; 1/2-  
 DR PDB; 1BHQ; X-ray; 1/2-  
 DR PDB; 1BDN; X-ray; 1/2-  
 DR PDB; 1IDO; X-ray; @=140-331.  
 DR PDB; 1JLM; X-ray; @=143-334.  
 DR PDB; 1JLM; X-ray; A=137-331.  
 DR PDB; 1MF7; X-ray; A=144-337.  
 DR PDB; 1N92; X-ray; A=140-335.  
 DR PDB; 1NA5; X-ray; A=144-345.  
 DR Genew; HGNC:6149; ITGAM.  
 DR MIM; 120980; -;  
 DR GO; GO:0008305; C:integrin complex; TAS.  
 DR GO; GO:0007155; P:cell adhesion; TAS.  
 DR InterPro; IPR000413; Integrin\_alpha.  
 DR InterPro; IPR02035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 3.  
 DR Pfam; PF00092; VWA; 1.  
 DR PRINTS; PRO1185; INTEGRIN.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00191; Int\_alpha; 5.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS50234; WFPA; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Receptor; Repeat; Signal;  
 KW Transmembrane.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1152 Integrin alpha-M.  
 Query 17 DLYEILDKSGSVL-HHNEIYFYVEQLAHKFIQSPQMSFVSTRTGTLTKLTD----- 71  
 Db 150 DIAFLIDGSGSIIPHDFRMRKEFVST-----VMEQLKSKTLFS-----LMQYSEBRRIH 199  
 Qy 72 -----REQIROGLEELQKVLPGCDTVMHEGRASQDYIENRGQYRTA-SVITALT 122  
 Db 200 FTPEKPNPNPNPSLVLPIQLL--GRTHATATGIRKVRRELFNTNGARKNAKFLIVIT 257  
 Qy 123 DGEIHELDLPFYSE--REANSRDIGAIYCGVVDVNETQLAR-----IADS--KDHVPP 173  
 Db 258 DGEKFGDPLGEBDYETPEADRE--GVIRVIYIGVDARFRSEKSRQELNTIASKEPRDHVQ 314  
 Qy 174 VNDGFQALOGIILHSILKSCIEILAEPTICAGESFQVAVVVGNGFPH 222  
 Db 315 VNN-FELKLTITQNLREK-----IPAIETQTGSSSSFEHMSQSGFSA 358  
 RESULT 14  
 Q8T5C2 PRELIMINARY; PRT; 453 AA.  
 AC Q8T5C2;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Proximal thread matrix protein 1.  
 OS Mytilus galloprovincialis (Mediterranean mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 OC Mytiloidea; Mytilidae; Mytilus.  
 OC NCBI\_TaxID=29158;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;  
 RA Sun C., Lucas J.M., Waite J.H.;  
 RT "Collagen-binding matrix proteins from elastomeric extraorganic  
 RL Biomacromolecules 3:1240-1248(2002).  
 DR EMBL; AY053391; AAL17974.1; -;

DR HSP, P20701, IMCN.  
DR GO:GO:0005198; F:structural molecule activity; IEA.  
DR InterPro: IPR02035; VWF\_A.  
DR Pfam: PF00092; VWA; 2.  
DR PRINTS: PRO0453; VWFADOMAIN.  
DR SMART: SM00327; VWA; 2.  
DR PROSITE: PS50234; VWFA; 2.  
DR Matrix protein.  
KW SEQUENCE 453 AA; 48784 MW; D604975C0C51E6D CRC64;  
SQ  
Query Match 9.5%; Score 145; DB 2; Length 453;  
Best Local Similarity 25.5%; Pred. No. 0.0044;  
Matches 51; Conservative 44; Mismatches 79; Indels 26; Gaps 12;  
OY 12 CYGGEFDYFIDKSGSV-----LHMNEIYFVEOLAHKF--ISFQ-LRMSFVSTGTT 64  
DB 248 CAGADIADFVDPASSIANNPNVQLMKNFMKQIVDRFKTGPDGQFAVTFADATK 307  
OY 65 LMKLTE--DRBOIRGLEELQKVLPG--GDTYMHGEFRASEQIYENRQ--YRTASV 117  
DB 308 QFGKDYSSKADIKGAIID---KVGPSIIQGTAGDGLNARLEV-FPRNCGGREYQKV 363  
OY 118 IIALTDGL--HEDLFYSREARSRDLGAIYCVG-KQNFQTGLARISDHPV 174  
DB 364 VILLTDGQNGHKS---PEHESSILRKEGVVAIVGVTGFLKSLINIASSEYV-T 418  
OY 175 NDGFQALQGIHSILKKSQI 194  
DB 419 TSFDFKSKIMEDVYKLACM 438  
RESULT 15  
ITAX HUMAN STANDARD; PRT; 1163 AA.  
AC P20702: O81VA6; 17, Created  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Integrin alpha-X precursor (Leukocyte adhesion glycoprotein p150,95  
alpha chain) (Leukocyte adhesion receptor p150,95) (CD11c) (Lew M5).  
GN Name=ITGAX; Synonyms=CD11C;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9816645; PubMed=3227687;  
RA Cordi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;  
RT "cDNA cloning and complete primary structure of the alpha subunit of a  
leukocyte adhesion glycoprotein, p150,95.";  
RL EMBO J. 6:4023-4028(1987).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90153906; PubMed=2303426;  
RA Cordi A.L., Garcia-Aguilar J., Springer T.A.;  
RT "Genomic structure of an integrin alpha subunit, the leukocyte p150,95  
molecule.";  
RL J. Biol. Chem. 265:2782-2788(1990).  
RN [3]  
RP ERDATUM.  
RA Cordi A.L., Garcia-Aguilar J., Springer T.A.;  
RL J. Biol. Chem. 265:12750-12751(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISUE=Blood;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Straube R.L., Feingold E.A., Groube L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F.,  
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,

RA Brownstein M.J., Ueda T.B., Toshiyuki S., Garnier P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Pahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywiński M.I., Skalska U., Smalins D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 20-43.  
RX MEDLINE=87167596; PubMed=3549901;  
RA Miller L.J., Wiebe M., Springer T.A.;  
RT "Purification and alpha subunit N-terminal sequences of human Mac-1  
and p150,95 leukocyte adhesion proteins.";  
RL J. Immunol. 138:2381-2383(1987).  
CC -1- FUNCTION: Integrin alpha-X/beta-2 is a receptor for fibrinogen. It  
recognizes the sequence G-P-R in fibrinogen. It mediates cell-cell  
interaction during inflammatory and chemotaxis.  
CC important in monocyte adhesion and chemotaxis.  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-X  
associates with beta-2.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and  
granulocytes.  
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins  
with I-domains do not undergo protease cleavage.  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -1- SIMILARITY: Contains 1 VWFA domain.  
CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD11c entry;  
WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cdlc.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL: M81695; AAA59180.1; -;  
DR EMBL: M29165; -; NOT ANNOTATED CDS.  
DR EMBL: M29487; AAA51620.1; ALT\_SEQ.  
DR EMBL: M29482; AAA51620.1; JOINED.  
DR EMBL: M29483; AAA51620.1; JOINED.  
DR EMBL: M29484; AAA51620.1; JOINED.  
DR EMBL: M29485; AAA51620.1; JOINED.  
DR EMBL: M29486; AAA51620.1; JOINED.  
DR EMBL: BC038237; AAA38237.1; -;  
DR PIR: A36584; RWHUIC.  
DR PDB: 1N3Y; X-ray; A=141-338.  
DR GeneW: HGNC:6152; ITGAX.  
DR MIM: 151510; -;  
DR GO:GO:0008105; C:integrin complex; TAS.  
DR GO:GO:0004872; F:receptor activity; TAS.  
DR GO:GO:0007155; P:cell adhesion; TAS.  
DR GO:GO:0009887; P:organogenesis; TAS.  
DR InterPro: IPR000413; Integrin\_alpha.  
DR InterPro: IPR002035; VWF\_A.  
DR Pfam: PF01839; FG-GAP; 3\_  
DR Pfam: PF00357; Integrin\_alpha; 1.  
DR Pfam: PF00092; VWA; 1.  
DR PRINTS: PRO1185; INTEGRINA.  
DR PRINTS: PRO0453; VWFADOMAIN.  
DR SMART: SM00191; Int\_alpha; 5.  
DR SMART: SM00327; VWA; 1\_  
DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.

DR PROSITE, PSS0234, VMPA, 1.  
 KW 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Polymorphism; Receptor; Repeat;  
 KW Signal; Transmembrane.

FT CHAIN 1 19 Integrin alpha-X.  
 FT DOMAIN 20 1163 Extracellular (Potential).  
 FT TRANSMEM 1108 1128 Potential.  
 FT DOMAIN 1129 1163 Cytoplasmic (Potential).  
 FT REPEAT 34 87 FG-GAP 1.  
 FT REPEAT 88 ? FG-GAP 2.  
 FT DOMAIN 165 351 VMPA.  
 FT REPEAT ? 401 FG-GAP 3.  
 FT REPEAT 402 453 FG-GAP 4.  
 FT REPEAT 455 517 FG-GAP 5.  
 FT REPEAT 518 576 FG-GAP 6.  
 FT REPEAT 581 633 FG-GAP 7.  
 FT CA\_BIND 466 474 Potential.  
 FT CA\_BIND 530 538 Potential.  
 FT CA\_BIND 593 601 Potential.  
 FT SITE 1131 1135 GFPKR motif.  
 FT DISULFID 69 76 By similarity.  
 FT DISULFID 108 126 By similarity.  
 FT DISULFID 655 712 By similarity.  
 FT DISULFID 771 777 By similarity.  
 FT DISULFID 848 863 By similarity.  
 FT DISULFID 998 1022 By similarity.  
 FT DISULFID 1027 1032 By similarity.  
 FT CARBOHYD 61 61 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 89 89 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 392 392 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 697 697 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 735 735 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 899 899 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 939 939 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1050 1050 N-linked (GlcNAc... ) (Potential).  
 FT VARIANT 48 48 W -> R (in dbSNP:11574633).  
 FT CONFLICT 209 209 /FTid=VAR\_018672.  
 FT CONFLICT 251 251 T -> S (in Ref. 4).  
 FT CONFLICT 469 469 T -> A (in Ref. 4).  
 FT CONFLICT 490 490 T -> S (in Ref. 4).  
 FT CONFLICT 547 547 G -> A (in Ref. 2).  
 FT CONFLICT 756 756 E -> K (in Ref. 4).  
 FT CONFLICT 819 819 D -> L (in Ref. 1).  
 FT CONFLICT 1163 1163 I -> V (in Ref. 4).  
 FT STRAND 150 157 SEK -> TPHPQDNV (in Ref. 4).  
 FT TURN 160 161  
 FT HELIX 164 178  
 FT TURN 179 180  
 FT TURN 183 185  
 FT STRAND 186 193  
 FT STRAND 197 201  
 FT HELIX 203 208  
 FT TURN 212 216  
 FT STRAND 226 226  
 FT HELIX 228 236  
 FT TURN 237 240  
 FT HELIX 242 244  
 FT TURN 245 245  
 FT TURN 248 249  
 FT STRAND 251 258  
 FT STRAND 263 263  
 FT HELIX 269 278  
 FT TURN 279 280  
 FT STRAND 282 288  
 FT HELIX 290 293  
 FT TURN 296 297  
 FT HELIX 298 304  
 FT HELIX 310 312  
 FT STRAND 313 316  
 FT HELIX 319 325

FT HELIX 326 334  
 FT TURN 335 335  
 SQ SEQUENCE 1163 AA; 127886 MW; 83658A13B5C5DE8F CRC64;  
 Query Match 9.4%; Score 143.5; DB 1; Length 1163;  
 Best local similarity 24.8%; Pred. No. 0.018; Indels 35; Gaps 10;  
 Matches 55; Conservative 43; Mismatches 89;  
 QY 17 DLVFIIDKSGSV-LHHNNEIYFVEQLAHKFIISPOLMSPFVFSFGTGTTLMLKLTEDRE-- 73  
 DB 151 DIVFLIDSSGSISSRNPFATMNFVRAVISQFQRPSTQFSLMQFENKQTHTFEERFRTS 210  
 QY 74 ---QIRGLELEQLVLLPEGDTYMEGPERASEQIYENRQGYRTAS-VTIALTQGEIHED 129  
 DB 211 NPLSLASVHQLQ-----GFTYATAIQNVVHRLFHASYGARRDATKILLIVITDGKEGD 265  
 QY 130 LFFYSERANRSRLGALIVCGV-----KDNENQLARIAD--SKDHVPVNDG 177  
 DB 266 SLDYKD-VTPMADAGIIRYALGVGLAPQENNSKEIAND---TASKPSQEHIFKVED- 318  
 QY 178 FOALQGIHSILKSCIEILAEPSITCAGESFOVVRGNGF 219  
 DB 319 FDALKDIGNQLEK-----IFAIGETFTSSSFLEMAQEGF 356

Search completed: June 13, 2005, 20:03:14  
 Job time : 99.3587 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 19:37:17 ; Search time 69.2347 Seconds  
(without alignments)  
1044.623 Million cell updates/sec

Title: us-09-970-076-2\_copy\_41\_227

Perfect score: 1 GGFDLYFLDKSGSVLHNM.....LGIHSILKKSCIEILAAE 187

Sequence: 1 GGFDLYFLDKSGSVLHNM.....LGIHSILKKSCIEILAAE 187

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneeqp19808:\*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	970	100.0	297	4	AAM38976 Human pol
2	970	100.0	328	7	Adi00558 Human TAN
3	970	100.0	328	7	Adm64584 Human TAN
4	970	100.0	333	5	Adm64584 Human TAN
5	970	100.0	333	3	Adm64584 Human TAN
6	970	100.0	333	7	Adm64584 Human TAN
7	970	100.0	333	7	Adm64584 Human TAN
8	970	100.0	333	7	Adm64584 Human TAN
9	970	100.0	342	7	Adm64584 Human TAN
10	970	100.0	345	7	Adm64584 Human TAN
11	970	100.0	345	7	Adm64584 Human TAN
12	970	100.0	368	5	Adm64584 Human TAN
13	970	100.0	384	5	Adm64584 Human TAN
14	970	100.0	403	4	Adm64584 Human TAN
15	970	100.0	403	4	Adm64584 Human TAN
16	970	100.0	403	8	Adm64584 Human TAN
17	970	100.0	460	7	Adm64584 Human TAN
18	970	100.0	479	7	Adm64584 Human TAN
19	970	100.0	504	7	Adm64584 Human TAN
20	970	100.0	529	7	Adm64584 Human TAN
21	970	100.0	540	7	Adm64584 Human TAN
22	970	100.0	540	7	Adm64584 Human TAN
23	970	100.0	549	7	Adm64584 Human TAN
24	970	100.0	549	7	Adm64584 Human TAN
25	970	100.0	549	7	Adm64584 Human TAN

26	970	100.0	549	7	Adm64590 TANGOL197
27	970	100.0	549	7	Adm64592 TANGOL197
28	970	100.0	551	7	Adi00550 Human TAN
29	970	100.0	551	7	Adm64576 Mouse TAN
30	970	100.0	564	5	Abb90750 Human Tum
31	970	100.0	564	5	Abb90724 Human Tum
32	970	100.0	564	5	Abp54904 Human ant
33	970	100.0	564	6	Abp54457 Human ant
34	970	100.0	564	6	Abu54431 Human tum
35	970	100.0	564	7	Adi00552 Human TAN
36	970	100.0	564	7	Adi00552 Human TAN
37	970	100.0	564	7	Adm64578 Human TAN
38	970	100.0	564	8	Adm64578 Human TAN
39	965	99.5	403	4	Aae01469 Human gen
40	965	99.5	403	4	Abg63873 Human alb
41	965	99.5	403	8	Adi00552 Human TAN
42	961	99.1	562	5	Abb90731 Mouse Tum
43	961	99.1	562	5	Abb90785 Mouse Tum
44	961	99.1	562	6	Abu54492 Mouse tum
45	961	99.1	562	6	Abu54438 Mouse tum

ALIGNMENTS

RESULT 1  
AAM38976 standard; protein: 297 AA.

22-OCT-2001 (first entry)  
Human polypeptide SEQ ID NO 2121.

Human, noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukaemia.

Homo sapiens.

WC20015312-A1.

26-JUL-2001.

26-DEC-2000; 2000MO-US034263.

23-DEC-1999; 99US-00471275.  
21-JAN-2000; 2000US-00488725.  
25-APR-2000; 2000US-00552317.  
20-JUN-2000; 2000US-00598042.  
19-JUL-2000; 2000US-00620312.  
03-AUG-2000; 2000US-00634450.  
14-SEP-2000; 2000US-00662191.  
19-OCT-2000; 2000US-00693036.  
29-NOV-2000; 2000US-00727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J, Zhao QA,  
Zhou F, Goodrich R, Drmanac RT.  
WPI; 2001-442253/47.  
N-PSDB; AAI58132.

Novel nucleic acids and polypeptides, useful for treating disorders such  
as central nervous system injuries.

Example 4; SEQ ID NO 2121; 10078BP; English.

```
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA038642-AA042213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 297 AA;
Query Match 100.0%; Score 970; DB 4; Length 297;
Best Local Similarity 100.0%; Pred. No. 1.6e-99;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGFPLVFLDKSGSVLHHNNEIYFVEQLAHKFTSPQLRMSFIYFSTRGTTLMKLTEDRE 60
DB 41 GGFPLVFLDKSGSVLHHNNEIYFVEQLAHKFTSPQLRMSFIYFSTRGTTLMKLTEDRE 100
QY 61 QIRGSELEOKVLPGGDTYHMEGFERRASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 120
DB 101 QIRGSELEOKVLPGGDTYHMEGFERRASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 160
QY 121 SERBANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSHLKKSC 180
DB 161 SERBANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSHLKKSC 220
QY 181 IEILAAE 187
DB 221 IEILAAE 227
RESULT 2
AD100558
ID AD100558 standard; protein; 328 AA.
XX
AC AD100558;
XX
DT 22-APR-2004 (first entry)
XX
DE Human TANGO 197 HisTag fusion protein - plasmid pO615.
XX
KW fusion; von Willebrand factor A-like domain; vWF; antibacterial;
KW cutaneous; inhalation anthrax; human; TANGO 197 HisTag fusion; mutant;
KW plasmid pO615; mutcin.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003144193-A1.
XX
PD 31-JUL-2003.
XX
PF 24-JUL-2002; 2002US-00201292.
XX
PR 20-DEC-2001; 2001US-00038307.
XX
PA (ROTT/) ROTTMAN J B.
PA (OKEE/) O'KEEFE T L.
PA (OZKA/) OZKAYNAK E.
PA (HEAL/) HEALEY J J.
XX
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;
XX
DR WPI; 2003-720708/68.
```

```
DR N-PSDB; AD100557.
XX
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous
PT to the vWF.
XX
XX Claim 45; SEQ ID NO 26; 86pp; English.
XX
XX The invention relates to a novel fusion polypeptide comprising a von
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino
CC acid sequence heterologous to the vWF. The polypeptide of the invention
CC demonstrates antibacterial activities whilst the composition and method
CC may be useful in preventing or ameliorating the symptoms of cutaneous
CC and/or inhalation anthrax. The current sequence is that of the human
CC TANGO 197 HisTag fusion protein of the invention.
XX
SQ Sequence 328 AA;
Query Match 100.0%; Score 970; DB 7; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.8e-99;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGFDLVFLDKSGSVLHHNNEIYFVEQLAHKFTSPQLRMSFIYFSTRGTTLMKLTEDRE 60
DB 41 GGFDLVFLDKSGSVLHHNNEIYFVEQLAHKFTSPQLRMSFIYFSTRGTTLMKLTEDRE 100
QY 61 QIRGSELEOKVLPGGDTYHMEGFERRASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 120
DB 101 QIRGSELEOKVLPGGDTYHMEGFERRASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 160
QY 121 SERBANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSHLKKSC 180
DB 161 SERBANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSHLKKSC 220
QY 181 IEILAAE 187
DB 221 IEILAAE 227
RESULT 3
ADM64584
ID ADM64584 standard; protein; 328 AA.
XX
AC ADM64584;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human TANGO197-His tag fusion protein #2.
XX
KW antibacterial; gene therapy;
KW von Willebrand factor A-like domain amino acid sequence;
KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;
KW inhalation anthrax; human; TANGO197; his tag; fusion protein.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US2003134786-A1.
XX
PD 17-JUL-2003.
XX
PF 20-DEC-2001; 2001US-00038307.
XX
PR 20-DEC-2001; 2001US-00038307.
XX
PA (ROTT/) ROTTMAN J B.
PA (OKEE/) O'KEEFE T L.
PA (OZKA/) OZKAYNAK E.
PA (HEAL/) HEALEY J J.
XX
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;
XX
```



DR WPI, 2003-829643/77.  
DR N-PSDB; ADM64583.  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
XX  
PS Claim 44; SEQ ID NO 26; 64pp; English.  
CC The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to *Bacillus anthracis*;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
CC sequence of a fusion protein comprising mature human TANGO197, thrombin  
CC cleavage site and his tag that can be used to treat exposure to or  
CC prevent a symptom of anthrax.  
XX  
SQ Sequence 328 AA;  
Query Match 100.0%; Score 970; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 1.8e-99;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGFPLVFLDKSGSVLHNMNIYYFVEQLAHKFTSPQLRMSFIVSTRGTTMLKLTEDRE 60  
DB 41 GGFPLVFLDKSGSVLHNMNIYYFVEQLAHKFTSPQLRMSFIVSTRGTTMLKLTEDRE 100  
QY QIRGGLBELQKVLPGGDTYHMEGFERASEQIYYENRQGYRTASVITALTDELHEDLFFY 120  
DB 101 QIRGGLBELQKVLPGGDTYHMEGFERASEQIYYENRQGYRTASVITALTDELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVCVGVKDFNETOLARIADSKDHFVPVNDGFQALOGIHSILKSC 180  
DB 161 SEREANSRDLGAIIVCVGVKDFNETOLARIADSKDHFVPVNDGFQALOGIHSILKSC 220  
QY 181 IEILAAE 187  
DB 221 IEILAAE 227  
RESULT 4  
AAB01422 standard; protein; 333 AA.  
ID AAB01422  
XX  
AC AAB01422;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Human TANGO 197.  
XX  
KM TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
KM graft versus-host diseases; rheumatoid arthritis; psoriasis;  
KM inflammatory bowel disease; septic shock; ulcerative colitis;  
KM Crohn's disease; chronic myelogenous leukemia; cancer; liver disease;  
KM Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia;  
KM autoimmune disease; myasthenia gravis; autoimmune diabetes;  
KM systemic lupus erythematosus; transgenic animal; diagnosis; prognosis;  
KM prophylactic; therapeutic; human.  
XX  
OS Homo sapiens.  
XX  
PN MO200039284-A1.  
XX  
PD 06-JUL-2000.  
XX  
PF 23-DEC-1999; 99MO-US031025.  
XX  
PR 30-DEC-1998; 98US-00223546.

XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Holtzman DA;  
XX  
XX WPI, 2000-465743/40.  
DR N-PSDB; AAA47455.  
XX  
XX  
PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224  
PT and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases.  
XX  
PS Claim 8; Fig 4; 209pp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents  
CC for regulating cellular processes like asthma, graft versus-host  
CC diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease,  
CC septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous  
CC leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis,  
CC Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis,  
CC autoimmune diabetes and systemic lupus erythematosus. The nucleic acids  
CC are also useful for producing transgenic animals and the TANGO  
CC polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239  
CC sequences are useful in forensic biology, for diagnostic assays,  
CC prognostic assays, pharmacogenomics and for monitoring clinical trials.  
CC TANGO polypeptides are suitable for both prophylactic and therapeutic  
CC methods for treating a subject at risk of a disorder or having a disorder  
CC associated with aberrant TANGO expression. A wide range of cellular  
CC disorders can be treated  
XX  
SQ Sequence 333 AA;  
Query Match 100.0%; Score 970; DB 3; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGFPLVFLDKSGSVLHNMNIYYFVEQLAHKFTSPQLRMSFIVSTRGTTMLKLTEDRE 60  
DB 41 GGFPLVFLDKSGSVLHNMNIYYFVEQLAHKFTSPQLRMSFIVSTRGTTMLKLTEDRE 100  
QY QIRGGLBELQKVLPGGDTYHMEGFERASEQIYYENRQGYRTASVITALTDELHEDLFFY 120  
DB 101 QIRGGLBELQKVLPGGDTYHMEGFERASEQIYYENRQGYRTASVITALTDELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVCVGVKDFNETOLARIADSKDHFVPVNDGFQALOGIHSILKSC 180  
DB 161 SEREANSRDLGAIIVCVGVKDFNETOLARIADSKDHFVPVNDGFQALOGIHSILKSC 220  
QY 181 IEILAAE 187  
DB 221 IEILAAE 227  
RESULT 5  
ABP54905 standard; protein; 333 AA.  
ID ABP54905  
XX  
AC ABP54905;  
XX  
DT 08-JAN-2003 (first entry)  
XX  
DE Human anthrax toxin receptor.  
XX  
KM Anthrax; toxin; receptor; human; antibacterial.  
XX  
OS Homo sapiens.  
XX  
OS  
XX  
XX Key Location/Qualifiers  
FH 1..27  
FT /label= Signal\_peptide  
FT 28..333  
FT Protein /label= Mature\_protein  
XX

PN WO200246228-A2.  
XX 13-JUN-2002.  
XX 03-OCT-2001; 2001WO-US030941.  
XX 05-DEC-2000; 2000US-0251481P.  
XX (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
PI WPI; 2002-713235/77.  
XX N-PSDB; ABV73883.  
XX  
XX Novel isolated polypeptide useful for identifying agent that prevents or  
PT reduces effect of anthrax toxin on host cell, for treating human or non-  
PT human animal suffering from anthrax.  
XX  
XX Claim 1; Page 42-43; 45pp; English.  
XX  
XX The present sequence is the protein sequence of a polypeptide identified  
CC as a human anthrax toxin receptor (ATR) polypeptide on the basis of  
CC identity to a newly isolated human ATR (see ABP54903). The 2 polypeptides  
CC are identical between amino acids 1-317, but differ thereafter at the C-  
CC terminus. The present polypeptide, encoded by IMAGE CLONE 4563020, has no  
CC previously known function, and there has been no prior indication that it  
CC is a complete or partial ATR. The invention provides ATR polypeptides and  
CC polynucleotides, vectors, host cells, and transgenic and knock-out  
CC animals. It also provides methods for identifying molecules that bind the  
CC ATR and which reduce the toxicity of anthrax toxin. A claimed method for  
CC treating anthrax in a human or animal involves administering an agent  
CC that inhibits binding between anthrax toxin protective antigen (PA) and  
CC ATR at a level effective to reduce the severity of anthrax. Suitable  
CC agents include the present polypeptide or a PA-binding fragment of it, a  
CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
CC nucleic acid  
XX  
XX Sequence 333 AA;  
SQ  
Query Match 100.0%; Score 970; DB 5; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGPDLFIIDKSGSVLHHMNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 60  
DB 41 GGGPDLFIIDKSGSVLHHMNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 100  
QY 61 QIRQGLELELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 120  
DB 101 QIRQGLELELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHISILKKSC 180  
DB 161 SEREANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHISILKKSC 220  
QY 181 IEIILAAE 187  
DB 221 IEIILAAE 227  
RESULT 6  
AD100534  
ID AD100534 standard; protein; 333 AA.  
XX  
XX AD100534;  
XX  
XX 22-APR-2004 (first entry)  
XX  
XX Human TANGO 197 protein.  
XX  
XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KW

KW cutaneous; inhalation anthrax; human; TANGO 197.  
XX  
XX Homo sapiens.  
XX US2003144193-A1.  
XX  
XX 31-JUL-2003.  
XX  
XX 24-JUL-2002; 2002US-00201292.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROTT/) ROTTMAN J B.  
XX (OKEE/) O'KEEFE T L.  
XX (OZKA/) OZKAYNAK E.  
XX (HEAL/) HEALEY J J.  
XX  
XX Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
PI WPI; 2003-720708/68.  
XX N-PSDB; AD100533.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
XX Claim 26; SEQ ID NO 2; 86pp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 protein of the invention.  
XX  
XX Sequence 333 AA;  
SQ  
Query Match 100.0%; Score 970; DB 7; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGGPDLFIIDKSGSVLHHMNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 60  
DB 41 GGGPDLFIIDKSGSVLHHMNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 100  
QY 61 QIRQGLELELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 120  
DB 101 QIRQGLELELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHISILKKSC 180  
DB 161 SEREANSRDLGAIIVYCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHISILKKSC 220  
QY 181 IEIILAAE 187  
DB 221 IEIILAAE 227  
RESULT 7  
ADM64568  
ID ADM64568 standard; protein; 333 AA.  
XX  
XX ADM64568;  
XX  
XX 03-JUN-2004 (first entry)  
XX  
XX Human von Willebrand factor A-like domain protein TANGO197.  
XX  
XX antibacterial; gene therapy;  
KW von Willebrand factor A-like domain amino acid sequence;  
KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
KW

KW inhalation anthrax; human; TANGO197.  
 XX Homo sapiens.  
 OS US2003134786-A1.  
 PN 17-JUL-2003.  
 XX 20-DEC-2001; 2001US-00038307.  
 PF 20-DEC-2001; 2001US-00038307.  
 PR 20-DEC-2001; 2001US-00038307.  
 XX (ROTT/) ROTTMAN J B.  
 PA (OKEE/) O'KEEFE T L.  
 PA (OZKA/) OZKAYNAK E.  
 PA (HEAL/) HEALEY J J.  
 XX Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
 PI WPI; 2003-829643/77.  
 DR N-PSDB; ADM64567.  
 XX  
 PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
 PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
 PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
 PT to the vWF.  
 PS  
 PS Claim 26; SEQ ID NO 2; 64pp; English.  
 CC The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the vWF. Also described are: a method of preventing or  
 CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to Bacillus anthracis;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
 CC sequence of a human von Willebrand factor A-like domain (vWF) amino acid  
 CC sequence TANGO197.  
 CC  
 SQ Sequence 333 AA;  
 Query Match 100.0%; Score 970; DB 7; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGFDLVFIIDKSGSVLHNNNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 60  
 DB 41 GGFDLVFIIDKSGSVLHNNNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 100  
 QY 61 QIRQGLEELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 120  
 DB 101 QIRQGLEELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 160  
 QY 121 SEREARSRLDGAIVYCVGVDFNETOLARIADSKDHVPVNDGFQALQGIHISILKSC 180  
 DB 161 SEREARSRLDGAIVYCVGVDFNETOLARIADSKDHVPVNDGFQALQGIHISILKSC 220  
 QY 181 IETLAEL 187  
 DB 221 IETLAEL 227  
 RESULT 8  
 ADI00554  
 ID ADI00554 standard; protein; 342 AA.  
 XX  
 AC ADI00554;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Human TANGO 197 FLAG fusion protein - plasmid p0613.  
 XX

KW fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
 KW cutaneous; inhalation anthrax; human; TANGO 197 FLAG fusion; mutant;  
 XX plasmid p0613; mutein.  
 OS Homo sapiens.  
 OS Synthetic.  
 OS US2003144193-A1.  
 PN 31-JUL-2003.  
 XX 24-JUL-2002; 2002US-00201292.  
 PF 24-JUL-2002; 2002US-00201292.  
 PR 20-DEC-2001; 2001US-00038307.  
 XX (ROTT/) ROTTMAN J B.  
 PA (OKEE/) O'KEEFE T L.  
 PA (OZKA/) OZKAYNAK E.  
 PA (HEAL/) HEALEY J J.  
 XX Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
 PI WPI; 2003-720708/68.  
 DR N-PSDB; ADI00553.  
 XX  
 PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
 PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
 PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
 PT to the vWF.  
 PS  
 PS Claim 45; SEQ ID NO 22; 86pp; English.  
 CC The invention relates to a novel fusion polypeptide comprising a von  
 CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
 CC acid sequence heterologous to the vWF. The polypeptide of the invention  
 CC demonstrates antibacterial activities whilst the composition and method  
 CC may be useful in preventing or ameliorating the symptoms of cutaneous  
 CC and/or inhalation anthrax. The current sequence is that of the human  
 CC TANGO 197 FLAG fusion protein of the invention.  
 CC  
 SQ Sequence 342 AA;  
 Query Match 100.0%; Score 970; DB 7; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGFDLVFIIDKSGSVLHNNNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 60  
 DB 50 GGFDLVFIIDKSGSVLHNNNEIYFVEQLAKHFIISPOLRMSFIYFSTRGTTLMKLTEDRE 109  
 QY 61 QIRQGLEELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 120  
 DB 110 QIRQGLEELQKVLPGGDTVMHGEFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFY 169  
 QY 121 SEREARSRLDGAIVYCVGVDFNETOLARIADSKDHVPVNDGFQALQGIHISILKSC 180  
 DB 170 SEREARSRLDGAIVYCVGVDFNETOLARIADSKDHVPVNDGFQALQGIHISILKSC 229  
 QY 181 IETLAEL 187  
 DB 230 IETLAEL 236  
 RESULT 9  
 ADM64580  
 ID ADM64580 standard; protein; 342 AA.  
 XX  
 AC ADM64580;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE Human TANGO197-FLAG epitope fusion protein.  
 XX

KM antibacterial; gene therapy;  
KM von Willebrand factor A-like domain amino acid sequence;  
KM vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
KM inhalation anthrax; human; TANGO197; FLAG; fusion protein; gene.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX US2003134786-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROTT/) ROTTMAN J B.  
XX (OKEE/) O'KEEFE T L.  
XX (OZKA/) OZKAYNAK E.  
XX (HEAL/) HEALEY J J.  
XX  
XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-829643/77.  
XX  
XX N-PSDB; ADM64579.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
XX ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
XX domain (vWF) amino acid sequence and an amino acid sequence heterologous  
XX to the vWF.  
XX  
XX  
XX Claim 44; SEQ ID NO 22; 64pp; English.  
XX  
XX The invention describes a fusion polypeptide comprising a von Willebrand  
XX factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
XX heterologous to the vWF. Also described are: a method of preventing or  
XX ameliorating a symptom of anthrax in a subject thought to be at risk for  
XX exposure to or suspected of having been exposed to Bacillus anthracis;  
XX and a pharmaceutical composition comprising the novel fusion polypeptide.  
XX The composition and method are useful in preventing or ameliorating  
XX symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
XX sequence of a fusion protein comprising human TANGO197 and FLAG that can  
XX be used to treat exposure to or prevent a symptom of anthrax.  
XX  
XX  
XX Sequence 342 AA;  
XX  
XX  
XX Query Match 100.0%; Score 970; DB 7; Length 342;  
XX Best Local Similarity 100.0%; Pred. No. 1.9e-99;  
XX Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX  
QY 1 GGFDFLFIIDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60  
DB 50 GGFDFLFIIDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 109  
QY 61 QIRGLEELQKVLPGGDTVMHGEFPERASEQIYYENRQGYRTASVIYIALTDGELHEDLFFY 120  
DB 110 QIRGLEELQKVLPGGDTVMHGEFPERASEQIYYENRQGYRTASVIYIALTDGELHEDLFFY 169  
QY 121 SEREANSRDLGAIYVCVGVDFNETOLARIADSKDHVFVNDGFQALQGIHISILKKSC 180  
DB 170 SEREANSRDLGAIYVCVGVDFNETOLARIADSKDHVFVNDGFQALQGIHISILKKSC 229  
QY 181 IEILAAE 187  
DB 230 IEILAAE 236

DT 22-APR-2004 (first entry)  
XX  
XX Human TANGO 197 HisTag fusion protein - plasmid p0614.  
DE  
XX  
XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KM cutaneous; inhalation anthrax; human; TANGO 197 HisTag fusion; mutant;  
KM plasmid p0614; mutein.  
XX  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX  
XX Key Location/Qualifiers  
XX  
XX Misc-difference 331  
XX /note= "Wild-type Cys replaced by Ser followed by  
XX thrombin cleavage site and His tag"  
XX  
XX US2003144193-A1.  
XX  
XX 31-JUL-2003.  
XX  
XX 24-JUL-2002; 2002US-00201292.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX  
XX (ROTT/) ROTTMAN J B.  
XX (OKEE/) O'KEEFE T L.  
XX (OZKA/) OZKAYNAK E.  
XX (HEAL/) HEALEY J J.  
XX  
XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-720708/68.  
XX  
XX N-PSDB; ADI00555.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
XX ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
XX domain (vWF) amino acid sequence and an amino acid sequence heterologous  
XX to the vWF.  
XX  
XX  
XX Claim 45; SEQ ID NO 24; 86pp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
XX Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
XX acid sequence heterologous to the vWF. The polypeptide of the invention  
XX demonstrates antibacterial activities whilst the composition and method  
XX may be useful in preventing or ameliorating the symptoms of cutaneous  
XX and/or inhalation anthrax. The current sequence is that of the human  
XX TANGO 197 HisTag fusion protein of the invention.  
XX  
XX  
XX Sequence 345 AA;  
XX  
XX  
XX Query Match 100.0%; Score 970; DB 7; Length 345;  
XX Best Local Similarity 100.0%; Pred. No. 2e-99;  
XX Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX  
QY 1 GGFDFLFIIDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60  
DB 41 GGFDFLFIIDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100  
QY 61 QIRGLEELQKVLPGGDTVMHGEFPERASEQIYYENRQGYRTASVIYIALTDGELHEDLFFY 120  
DB 101 QIRGLEELQKVLPGGDTVMHGEFPERASEQIYYENRQGYRTASVIYIALTDGELHEDLFFY 160  
QY 121 SEREANSRDLGAIYVCVGVDFNETOLARIADSKDHVFVNDGFQALQGIHISILKKSC 180  
DB 161 SEREANSRDLGAIYVCVGVDFNETOLARIADSKDHVFVNDGFQALQGIHISILKKSC 220  
QY 181 IEILAAE 187  
DB 221 IEILAAE 227

RESULT 11

AD	M64582
ID	ADM64582 standard; protein; 345 AA.
AC	ADM64582;
XX	
DT	03-JUN-2004 (first entry)
XX	
DE	Human TANGO197-His tag fusion protein #1.
XX	
KW	antibacterial; gene therapy;
KV	von Willebrand factor A-like domain amino acid sequence;
KM	vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;
XX	inhalation anthrax; human; TANGO197; his tag; fusion protein; gene.
OS	Homo sapiens.
XX	Synthetic.
PN	US2003134786-A1.
PD	17-JUL-2003.
PF	20-DEC-2001; 2001US-00038307.
PR	20-DEC-2001; 2001US-00038307.
PA	(ROTT/) ROTTMAN J B.
PA	(OKEE/) O'KEEFE T L.
PA	(OZKA/) OZKAYNAK E.
PA	(HEAL/) HEALEY J J.
P1	Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;
DR	MP1; 2003-829643/77.
XX	N-Psdb; ADM64581.
PT	New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or
PT	ameliorating symptoms of anthrax comprises a von Willebrand factor A-like
PT	domain (vWF) amino acid sequence and an amino acid sequence heterologous
XX	to the vWF.
PS	Claim 44; SEQ ID NO 24; 64bp; English.
XX	
CC	The invention describes a fusion polypeptide comprising a von Willebrand
CC	factor A-like domain (vWF) amino acid sequence and an amino acid sequence
CC	heterologous to the vWF. Also described are: a method of preventing or
CC	ameliorating a symptom of anthrax in a subject thought to be at risk for
CC	exposure to or suspected of having been exposed to Bacillus anthracis;
CC	and a pharmaceutical composition comprising the novel fusion polypeptide.
CC	The composition and method are useful in preventing or ameliorating
CC	symptoms of cutaneous and/or inhalation anthrax. This is the amino acid
CC	sequence of a fusion protein comprising human TANGO197, thrombin cleavage
CC	site and his tag that can be used to treat exposure to or prevent a
XX	symptom of anthrax.
XX	
SO	Sequence 345 AA;
QY	Query Match 100.0%; Score 970; DB 7; Length 345;
Dd	Best Local Similarity 100.0%; Pred. No. 2e-99; Mismatches 0; Indels 0; Gaps 0
Matches	187; Conservative 0;
QY	1 GGFDLYFLIDSGSVLAHMHNEIYYVEQLAHKFISPOLMSFIVSTRGTTLMKLTEDRE 60
Dd	41 GGFDLYFLIDSGSVLAHMHNEIYYVEQLAHKFISPOLMSFIVSTRGTTLMKLTEDRE 100
QY	61 QIRQGLEELQKYLPGGDITVMHEGFERRASFOIYYENRQGRTASVIATLTDGELHEDLFPE 120
Dd	101 QIRQGLEELQKYLPGGDITVMHEGFERRASFOIYYENRQGRTASVIATLTDGELHEDLFPE 160
QY	121 SERANRRBDIGAIYYCVGVNDPENTQTARINDSKDHVPVNDGFOALOGIIHSILTKSC 180
Dd	161 SERANRRBDIGAIYYCVGVNDPENTQTARINDSKDHVPVNDGFOALOGIIHSILTKSC 220
QY	181 IEILLAE 187

DB	221	IRILAB	227
RESULT 12			
ABP54903			
ID	ABP54903	standard; protein; 368 AA.	
AC	ABP54903;		
XX			
XX	08-JAN-2003	(first entry)	
DT			
XX			
DE	Human anthrax toxin receptor.		
XX			
XX	Anthrax; toxin; receptor; human; antibacterial.		
KW			
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	Peptide	1..27	
FT		/label= signal_peptide	
FT	Region	27..321	
FT		/note= "PP-binding fragment, begins at any amino acid in the range 27-43 and ends at any amino acid in the range 221-321, region specifically described in Claim 3"	
FT		28..368	
FT	Protein	/label= Mature_protein	
FT		28..320	
FT	Domain	/note= "extracellular domain"	
FT		44..216	
FT	Domain	/note= "von Willebrand factor A domain"	
FT		50	
FT	Region	/note= "Forms metal ion-dependent adhesion site (MIDAS) motif with amino acid residues 52, 54, 118 and 150"	
FT		52	
FT	Region	/note= "forms metal ion-dependent adhesion site (MIDAS) motif with amino acid residues 50, 54, 118 and 150"	
FT		54	
FT	Region	/note= "Forms metal ion-dependent adhesion site (MIDAS) motif with amino acid residues 50, 52, 118 and 150"	
FT		118	
FT	Region	/note= "forms metal ion-dependent adhesion site (MIDAS) motif with amino acid residues 50, 52, 54 and 150"	
FT		150	
FT	Region	/note= "forms metal ion-dependent adhesion site (MIDAS) motif with amino acid residues 50, 52, 54 and 118"	
FT		320..343	
FT	Domain	/note= "putative transmembrane domain"	
FT		344..368	
FT	Domain	/note= "cytoplasmic domain"	
XX			
PN	WO200246228-A2.		
XX			
PD	13-JUN-2002.		
XX			
PF	03-OCT-2001; 2001WO-US030941.		
XX			
PR	05-DEC-2000; 2000US-0251481P.		
XX			
PA	(WISC ) WISCONSIN ALUMNI RES FOUND.		
PI	Young JAT, Bradley KA, Collier RJ, Mogridge JS;		
XX			
DR	WPI; 2002-713235/77.		
XX			
DR	N-PSDB; ABV73881.		
XX			
XX	Novel isolated polypeptide useful for identifying agent that prevents or reduces effect of anthrax toxin on host cell, for treating human or non-human animal suffering from anthrax.		
XX			
XX	Claim 1; Page 29-30; 45pp; English.		
XX			
XX	The present sequence is the protein sequence of a human surface-bound		

CC anthrax toxin receptor (ATR), as predicted from an isolated cDNA clone.  
 CC Anthrax toxin protective antigen (PA) binds to the ATR at a von  
 CC Willebrand factor A domain located in the extracellular domain of ATR.  
 CC The invention provides ATR polypeptides and polynucleotides, vectors,  
 CC host cells, and transgenic and knock-out animals. It also provides the  
 CC methods for identifying molecules that bind the ATR and which reduce the  
 CC toxicity of anthrax toxin. A claimed method for treating anthrax in a  
 CC human or animal involves administering an agent that inhibits binding  
 CC between PA and ATR at a level effective to reduce the severity of  
 CC anthrax. Suitable agents include ATR or a PA-binding fragment of ATR, a  
 CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
 CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
 CC nucleic acid  
 CC  
 XX Sequence 368 AA;  
 SQ  
 Query Match 100.0%; Score 970; DB 5; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-99;  
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGFDFYFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFVFSRGTTLTKLTEDRE 60  
 Db 41 GGFDFYFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFVFSRGTTLTKLTEDRE 100  
 QY 61 QIRQGLEELQKVLPGGDTYHMEGFERASEQIYENRQGYRTASVYIALTDGEIHEDLFFY 120  
 Db 101 QIRQGLEELQKVLPGGDTYHMEGFERASEQIYENRQGYRTASVYIALTDGEIHEDLFFY 160  
 QY 121 SEREANSRDLGAIYVCVGVDPFNETOLARIADSKDHFVPNDGFOALQGIHSLKSC 180  
 Db 161 SEREANSRDLGAIYVCVGVDPFNETOLARIADSKDHFVPNDGFOALQGIHSLKSC 220  
 QY 181 IEILAAE 187  
 Db 221 IEILAAE 227  
 Db  
 RESULT 13  
 ADM64586  
 ID ADM64586 standard; protein; 384 AA.  
 AC ADM64586;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE TANGO197 extracellular domain-mutant IgG Fc fusion protein #1.  
 XX  
 KW antibacterial; gene therapy;  
 KW von Willebrand factor A-like domain amino acid sequence;  
 KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
 KW inhalation anthrax; human; TANGO197; immunoglobulin G; IgG;  
 KW fragment of crystallisation; Fc.  
 KW  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX US2003134786-A1.  
 PN  
 PD 17-JUL-2003.  
 XX  
 PF 20-DEC-2001; 2001US-00038307.  
 XX  
 PR 20-DEC-2001; 2001US-00038307.  
 XX  
 XX (ROTT/) ROTTMAN J B.  
 PA (OKEE/) O'KEEFE T L.  
 PA (OZKA/) OZKAYNAK E.  
 PA (HEAL/) HEALEY J J.  
 PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
 XX WPI; 2003-829643/77.  
 DR N-PSDB; ADM64585.  
 DR

XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
 PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
 PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
 PT to the vWF.  
 XX  
 PS Claim 44; SEQ ID NO 10; 64pp; English.  
 XX  
 CC The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the vWF. Also described are: a method of preventing or  
 CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to Bacillus anthracis;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
 CC sequence of a fusion protein comprising human TANGO197 extracellular  
 CC region, minus the DG residues closest to the transmembrane region, and  
 CC immunoglobulin G (196) fragment of crystallisation (Fc) with mutations  
 CC L235A and G237A.  
 CC  
 XX Sequence 384 AA;  
 SQ  
 Query Match 100.0%; Score 970; DB 7; Length 384;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-99;  
 Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGFDFYFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFVFSRGTTLTKLTEDRE 60  
 Db 41 GGFDFYFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFVFSRGTTLTKLTEDRE 100  
 QY 61 QIRQGLEELQKVLPGGDTYHMEGFERASEQIYENRQGYRTASVYIALTDGEIHEDLFFY 120  
 Db 101 QIRQGLEELQKVLPGGDTYHMEGFERASEQIYENRQGYRTASVYIALTDGEIHEDLFFY 160  
 QY 121 SEREANSRDLGAIYVCVGVDPFNETOLARIADSKDHFVPNDGFOALQGIHSLKSC 180  
 Db 161 SEREANSRDLGAIYVCVGVDPFNETOLARIADSKDHFVPNDGFOALQGIHSLKSC 220  
 QY 181 IEILAAE 187  
 Db 221 IEILAAE 227  
 Db  
 RESULT 14  
 AAE01439  
 ID AAE01439 standard; protein; 403 AA.  
 AC AAE01439;  
 XX  
 DT 17-JUL-2001 (first entry)  
 XX  
 DE Human gene 4 encoded secreted protein HMLPR02, SEQ ID NO:94.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification;  
 KW chromosome 19.  
 KW  
 OS Homo sapiens.  
 OS  
 OS  
 FH Key  
 FT Peptide 1..27  
 FT Protein /label= signal\_peptide  
 FT 28..403  
 FT /note= "Mature human secreted protein"

XX WO200134626-A1.  
XX 17-MAY-2001.  
XX 01-NOV-2000; 2000WO-US030045.  
XX 05-NOV-1999; 99US-0163581P.  
XX 30-JUN-2000; 2000US-0215133P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Komatsu GA, Moore PA, Birse CE, Ni J;  
XX WPI; 2001-308778/32.  
XX N-PSDB; AAD05303.  
XX  
XX New nucleic acid molecules encoding 28 human secreted proteins for  
XX diagnosing, preventing, treating or ameliorating medical conditions and  
XX used as food additives or preservatives.  
XX  
XX Claim 11; Page 485-486; 562pp; English.  
XX  
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
XX protein genes, and AAB01436-AAB01513 represent the proteins they encode.  
XX AAB01514-AAB01544 represent human secreted protein fragments or variants.  
XX The genes and their secreted proteins are useful for preventing, treating  
XX or ameliorating medical conditions, e.g., by protein or gene therapy.  
XX Pathological conditions can be diagnosed by determining the amount of the  
XX new protein in a sample or by determining the presence of mutations in  
XX the new genes. Specific uses are described for each of the 28 genes,  
XX based on the tissues in which they are most highly expressed, and include  
XX developing products for the diagnosis or treatment of proliferative  
XX disorders, cancer, tumours, foetal and developmental abnormalities,  
XX haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
XX diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
XX neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
XX cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
XX psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
XX angiotensin-related disorders, kidney disorders, gastrointestinal disorders,  
XX pregnancy-related disorders, endocrine disorders, and infections. The  
XX proteins can also be used to aid wound healing and epithelial cell  
XX proliferation, to prevent skin aging due to sunburn, to maintain organs  
XX before transplantation, for supporting cell culture of primary tissues,  
XX to regenerate tissues, to identify their cognate ligands or binding  
XX partners, and in chemotaxis, and can be used as a food additive or  
XX preservative to modify storage properties. Antibodies specific for a  
XX protein of the invention can be used in alleviating symptoms associated  
XX with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
XX radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
XX present sequence represents a human secreted protein of the invention  
XX  
XX Sequence 403 AA;  
XX  
XX Query Match 100.0%; Score 970; DB 4; Length 403;  
XX Best Local Similarity 100.0%; Pred. No. 2,4e-99;  
XX Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
Qy 1 GGFPLVFLDKSGSVLHNMNIYFVEQLAKFISPOLRMSFIYFSTRTGLMLKLTEDRE 60  
Db 41 GGFPLVFLDKSGSVLHNMNIYFVEQLAKFISPOLRMSFIYFSTRTGLMLKLTEDRE 100  
Qy QIRGLELELOKVLPGGDTYMHGFERASEQIYYENRGYRTASVIALTDGELHEDLPFY 120  
Db 101 QIRGLELELOKVLPGGDTYMHGFERASEQIYYENRGYRTASVIALTDGELHEDLPFY 160  
Qy 121 SERBANRSDIGAIVYCVGVDFNETQARIADSKDHFVYVNDGFOALQGIHSILKSC 180  
Db 161 SERBANRSDIGAIVYCVGVDFNETQARIADSKDHFVYVNDGFOALQGIHSILKSC 220  
Qy 181 IETLAE 187  
Db 221 IETLAE 227

RESULT 15  
ABG63874  
ID ABG63874 standard; protein; 403 AA.  
XX  
XX ABG63874;  
XX  
XX 27-AUG-2002 (first entry)  
XX  
XX Human albumin fusion protein #549.  
XX  
XX Albumin fusion protein, therapeutic protein X; human albumin; HA;  
XX human serum albumin; HSA; cancer; reproductive disorder;  
XX digestive disorder; immune disorder; endocrine disorder;  
XX haematopoietic disorder; neural disorder; connective disorder;  
XX cytosolic; antifertility; antiinflammatory; anticancer;  
XX immunomodulator; anti-HIV; antidiabetic; hemostatic; neurotropic;  
XX neuroprotective; antiParkinsonian; antimicrobial; neuroleptic;  
XX osteopathic; antiarthritic.  
XX  
XX Homo sapiens.  
XX Synthetic.  
XX  
XX WO200177137-A1.  
XX  
XX 18-OCT-2001.  
XX  
XX 12-APR-2001; 2001WO-US011988.  
XX  
XX 12-APR-2000; 2000US-0229358P.  
XX 25-APR-2000; 2000US-0199384P.  
XX 21-DEC-2000; 2000US-0256931P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Haseltine WA;  
XX WPI; 2002-010886/01.  
XX  
XX New fusion protein for treating disease e.g. diabetes comprises an  
XX albumin fused to a therapeutic protein.  
XX  
XX Claim 1; Page 874-875; 2102pp; English.  
XX  
XX The present invention relates to albumin fusion proteins comprising a  
XX therapeutic protein X and human albumin (HA, also known as human serum  
XX albumin, HSA). The proteins are useful for treating a disease or disorder  
XX that may be modulated by therapeutic protein X. The albumin extends the  
XX shelf-life of protein X, and may increase its biological in vitro/in vivo  
XX activity. The protein is useful for treating and diagnosing disorders  
XX such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
XX disease, ulcerative colitis), immune disorders (e.g. acquired  
XX immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
XX haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
XX Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
XX schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
XX ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX  
XX Sequence 403 AA;  
XX  
XX Query Match 100.0%; Score 970; DB 5; Length 403;  
XX Best Local Similarity 100.0%; Pred. No. 2,4e-99;  
XX Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
Qy 1 GGFPLVFLDKSGSVLHNMNIYFVEQLAKFISPOLRMSFIYFSTRTGLMLKLTEDRE 60  
Db 41 GGFPLVFLDKSGSVLHNMNIYFVEQLAKFISPOLRMSFIYFSTRTGLMLKLTEDRE 100  
Qy QIRGLELELOKVLPGGDTYMHGFERASEQIYYENRGYRTASVIALTDGELHEDLPFY 120  
Db 101 QIRGLELELOKVLPGGDTYMHGFERASEQIYYENRGYRTASVIALTDGELHEDLPFY 160

Qy	121	SEREANRSRDIGAIVCVGVKDFNETQIARIADSKDHVFPVNDGFQALOGIIHSILKSC	180
Db	161	SEREANRSRDIGAIVCVGVKDFNETQIARIADSKDHVFPVNDGFQALOGIIHSILKSC	220
Qy	181	IEILAAE	187
Db	221	IEILAAE	227

Search completed: June 13, 2005, 19:56:52  
Job time : 71.2347 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using SW model

Run on: June 13, 2005, 19:46:02 ; Search time 18.054 Seconds  
(without alignments)  
773.200 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_41\_227

Perfect score: 1 GGFDPYILDKSGSVLHRM.....LQGIHSILKSCIEILAAE 187

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	142.5	14.7	435	5 PCT-US95-04439-1	Sequence 1, Appl
2	141.5	14.6	1151	1 US-08-286-889-37	Sequence 37, Appl
3	141.5	14.6	1151	1 US-08-485-618-37	Sequence 37, Appl
4	141.5	14.6	1151	1 US-08-362-652-37	Sequence 37, Appl
5	141.5	14.6	1151	2 US-08-605-672-37	Sequence 37, Appl
6	141.5	14.6	1151	2 US-08-482-293A-37	Sequence 37, Appl
7	141.5	14.6	1151	3 US-08-943-363-37	Sequence 37, Appl
8	141.5	14.6	1151	3 US-09-193-043-37	Sequence 37, Appl
9	141.5	14.6	1151	4 US-09-688-307A-37	Sequence 37, Appl
10	141.5	14.6	1151	4 US-09-350-259-37	Sequence 37, Appl
11	141.5	14.6	1161	1 US-08-485-618-55	Sequence 55, Appl
12	141.5	14.6	1161	1 US-08-362-652-55	Sequence 55, Appl
13	141.5	14.6	1161	2 US-08-605-672-55	Sequence 55, Appl
14	141.5	14.6	1161	2 US-08-482-293A-55	Sequence 55, Appl
15	141.5	14.6	1161	2 US-08-943-363-55	Sequence 55, Appl
16	141.5	14.6	1161	3 US-09-193-043-55	Sequence 55, Appl
17	141.5	14.6	1161	4 US-09-688-307A-55	Sequence 55, Appl
18	141.5	14.6	1161	4 US-09-350-259-55	Sequence 55, Appl
19	138.5	14.3	1155	1 US-08-286-889-46	Sequence 46, Appl
20	138.5	14.3	1155	1 US-08-485-618-46	Sequence 46, Appl
21	138.5	14.3	1155	2 US-08-362-652-46	Sequence 46, Appl
22	138.5	14.3	1155	2 US-08-605-672-46	Sequence 46, Appl
23	138.5	14.3	1155	2 US-08-482-293A-46	Sequence 46, Appl
24	138.5	14.3	1155	2 US-08-943-363-46	Sequence 46, Appl
25	138.5	14.3	1155	3 US-09-193-043-46	Sequence 46, Appl
26	138.5	14.3	1155	4 US-09-688-307A-46	Sequence 46, Appl
27	138.5	14.3	1155	4 US-09-350-259-46	Sequence 46, Appl

28	138.5	14.3	1161	1 US-08-485-618-53	Sequence 53, Appl
29	138.5	14.3	1161	1 US-08-362-652-53	Sequence 53, Appl
30	138.5	14.3	1161	2 US-08-605-672-53	Sequence 53, Appl
31	138.5	14.3	1161	2 US-08-482-293A-53	Sequence 53, Appl
32	138.5	14.3	1161	2 US-08-943-363-53	Sequence 53, Appl
33	138.5	14.3	1161	3 US-09-193-043-53	Sequence 53, Appl
34	138.5	14.3	1161	4 US-09-688-307A-53	Sequence 53, Appl
35	138.5	14.3	1161	4 US-09-350-259-53	Sequence 53, Appl
36	136.5	14.1	216	4 US-09-795-872-5	Sequence 43, Appl
37	132.5	13.7	1152	2 US-08-476-062A-43	Sequence 43, Appl
38	132.5	13.7	1152	5 PCT-US96-01314-43	Sequence 43, Appl
39	132.5	13.7	1152	6 5424399-2	Patent No. 5424399
40	132.5	13.7	1152	6 5424399-2	Patent No. 5424399
41	132.5	13.7	1153	1 US-08-173-497-3	Sequence 3, Appl1
42	132.5	13.7	1153	1 US-08-286-889-3	Sequence 3, Appl1
43	132.5	13.7	1153	1 US-08-485-618-3	Sequence 3, Appl1
44	132.5	13.7	1153	1 US-08-362-652-3	Sequence 3, Appl1
45	132.5	13.7	1153	2 US-08-605-672-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
PCT-US95-04439-1  
Sequence 1, Application PC/TUS9504439  
GENERAL INFORMATION:  
APPLICANT: The Upjohn Company  
APPLICANT: Heinrichson, Robert L.  
APPLICANT: Anderson, Donald C.  
APPLICANT: Tomlich, Che-Shen C.  
APPLICANT: Fairbanks, Michael B.  
APPLICANT: Baile, Mary L.  
TITLE OF INVENTION: MAC-1 I-DOMAIN PROTEIN USEFUL IN  
BLOCKING ADHESION AND MIGRATION OF NEUTROPHILS  
TITLE OF INVENTION: BLOCKING ADHESION AND MIGRATION OF NEUTROPHILS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: The Upjohn Company, Intellectual Property Law  
STREET: 301 Henrietta  
CITY: Kalamazoo  
STATE: MI  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Gateway 2000 P5-90  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04439  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Darnley, James D., Jr.  
REGISTRATION NUMBER: 33,673  
REFERENCE/DOCKET NUMBER: 4767.P CNI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616/385-5210  
TELEFAX: 616/385-6897  
TELEX: 224401  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04439-1

Query Match 14.7%; Score 142.5; DB 5; Length 435;  
Best Local Similarity 26.94; Pred. No. 8.4e-08;  
Matches 54; Conservative 41; Mismatches 67; Indels 39; Gaps 11;



Db 257 PLEYSVDVIPADKX---GIIRYALGVGDAPQEPFALKEINTIGSAPPQDHVFKVGN-FAA 312  
QY 168 LOGIHSILKK 178  
Db 313 LRSTQRQLQEK 323

RESULT 4  
US-08-362-652-37  
Sequence 37, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-37

Query Match 14.6%; Score 141.5; DB 1; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
QY 4 DLVFIIDKGSV-LHNMNIYVVEQLAKFISPOLMSFIVSTGTTLMKTE-----57  
Db 142 DIAFLIDSGSSINQRPFAWKDFVKALMGEPASTSTLFSIMQYSNLTKEHFTFTEPKNTL 201  
QY 58 DREQIQGLELEQKVLPGSDTYMHGEPERASQIYYENRGYRTA-SVIALTDGELHED 116  
Db 202 DPQSLVDPIVQLQ-----GLTYTATGIRTYMEELFHSKSGRSKSAKKILVITDQKTRD 256  
QY 117 LEPYSE--BEANSRDLGAIYVCVGYD-FNE-TOLARI-----ADSKDHVFPVNDGFOA 167  
Db 257 PLEYSVDVIPADKX---GIIRYALGVGDAPQEPFALKEINTIGSAPPQDHVFKVGN-FAA 312  
QY 168 LOGIHSILKK 178  
Db 313 LRSTQRQLQEK 323

RESULT 5  
US-08-605-672-37  
Sequence 37, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-37

Query Match 14.6%; Score 141.5; DB 2; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
QY 4 DLVFIIDKGSV-LHNMNIYVVEQLAKFISPOLMSFIVSTGTTLMKTE-----57  
Db 142 DIAFLIDSGSSINQRPFAWKDFVKALMGEPASTSTLFSIMQYSNLTKEHFTFTEPKNTL 201  
QY 58 DREQIQGLELEQKVLPGSDTYMHGEPERASQIYYENRGYRTA-SVIALTDGELHED 116  
Db 202 DPQSLVDPIVQLQ-----GLTYTATGIRTYMEELFHSKSGRSKSAKKILVITDQKTRD 256  
QY 117 LEPYSE--BEANSRDLGAIYVCVGYD-FNE-TOLARI-----ADSKDHVFPVNDGFOA 167  
Db 257 PLEYSVDVIPADKX---GIIRYALGVGDAPQEPFALKEINTIGSAPPQDHVFKVGN-FAA 312  
QY 168 LOGIHSILKK 178  
Db 313 LRSTQRQLQEK 323

```
RESULT 6
US-08-482-293A-37
; Sequence 37, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; TITLE OF INVENTION: Van der Vlieten, Monica
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Seer Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-293A-37

Query Match 14.6%; Score 141.5; DB 2; Length 1151;
Best Local Similarity 27.7%; Pred. NO. 4.5e-07;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

QY 4 DLYFIIDKSGSV-LHHNNEIYFVEQLAHKEISPOLRMSFIVSTRGTTLMKLT----- 57
DB 142 DIATLIDSSGSIQRDPQKMDVFKALMGFFASTSTLFSLMQYSNLIKHTFTFEFNKIL 201
QY 58 DREQIRGLLELQKVLPGSDTYMHEGFERASEQIYYENRQGYRTA-SVITALTGDELHED 116
DB 202 DPQSLVDPIVQLQ-----GLTYTATGIRTYMEELFHSKNSRSKAKILLVITDGQKXRD 256
QY 117 LFFYSE--REANRSDIGAIVYCVKVD-FNE-TQLARI-----ADSKDHVFPVNDGFOA 167
DB 257 PLEYSVDVIPADKA---GIIRYALGVDAFOEPTALKEINTIGSAPPQDHVFKVGN-FAA 312
QY 168 LOGIHSILKK 178
DB 313 LRSTORQLOEK 323

RESULT 7
US-08-943-363-37
```

```
; Sequence 37, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; TITLE OF INVENTION: Van der Vlieten, Monica
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Seer Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-37

Query Match 14.6%; Score 141.5; DB 2; Length 1151;
Best Local Similarity 27.7%; Pred. NO. 4.5e-07;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

QY 4 DLYFIIDKSGSV-LHHNNEIYFVEQLAHKEISPOLRMSFIVSTRGTTLMKLT----- 57
DB 142 DIATLIDSSGSIQRDPQKMDVFKALMGFFASTSTLFSLMQYSNLIKHTFTFEFNKIL 201
QY 58 DREQIRGLLELQKVLPGSDTYMHEGFERASEQIYYENRQGYRTA-SVITALTGDELHED 116
DB 202 DPQSLVDPIVQLQ-----GLTYTATGIRTYMEELFHSKNSRSKAKILLVITDGQKXRD 256
QY 117 LFFYSE--REANRSDIGAIVYCVKVD-FNE-TQLARI-----ADSKDHVFPVNDGFOA 167
DB 257 PLEYSVDVIPADKA---GIIRYALGVDAFOEPTALKEINTIGSAPPQDHVFKVGN-FAA 312
QY 168 LOGIHSILKK 178
DB 313 LRSTORQLOEK 323

RESULT 8
US-09-193-043-37
; Sequence 37, Application US/09193043
; Patent No. 6251395
```

GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 6251395e1 Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/193,043  
CURRENT FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 37  
LENGTH: 1151  
TYPE: PRT  
ORGANISM: Rattus rattus  
US-09-193-043-37

Query Match 14.6%; Score 141.5; DB 3; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 4 DLYFLDKSGSV-LHNNIEIYFVEQLAKKFIQPOLMSPIVSTRTGLMKLE----- 57  
Db 142 DIAFLDSSGSSINORDPAQKMDFYKALMGEPASTSTFLMQYSLNLIKHTFTFEPKNIL 201  
Qy 58 DREGIROGLBELQVLPFGDTYMEGPERASEQIYYENRQGYRTA-SVIALTDGLHED 116  
Db 202 DPGSLVPIYQLQ-----GLTYATGIRTYMBELFHSKNGSRKSAKKILLVITDGGCKRD 256  
Qy 117 LFPYSE--REANRSDGAIYVCYVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 167  
Db 257 PLEYSDVTPADAKA---GIIRYALGVDAFOEPTALKEINTIGSAPPDHFVKGN-FAA 312  
Qy 168 LOGIHSILKK 178  
Db 313 LRSIQRLQEK 323

RESULT 9  
US-09-688-307A-37  
Sequence 37, Application US/09688307A  
Patent No. 6432404  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 6432404e1 Human Beta-2  
FILE REFERENCE: 27866/36646  
CURRENT APPLICATION NUMBER: US/09/688,307A  
CURRENT FILING DATE: 2000-10-13  
EARLIER APPLICATION NUMBER: 09/193,043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/605,672  
EARLIER FILING DATE: 1996-02-22  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 37  
LENGTH: 1151  
TYPE: PRT  
ORGANISM: Rattus rattus

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 361  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 464  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 486  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 506  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1117  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1118  
OTHER INFORMATION: Xaa = any or unknown amino acid  
US-09-688-307A-37

Query Match 14.6%; Score 141.5; DB 4; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 4 DLYFLDKSGSV-LHNNIEIYFVEQLAKKFIQPOLMSPIVSTRTGLMKLE----- 57  
Db 142 DIAFLDSSGSSINORDPAQKMDFYKALMGEPASTSTFLMQYSLNLIKHTFTFEPKNIL 201  
Qy 58 DREGIROGLBELQVLPFGDTYMEGPERASEQIYYENRQGYRTA-SVIALTDGLHED 116  
Db 202 DPGSLVPIYQLQ-----GLTYATGIRTYMBELFHSKNGSRKSAKKILLVITDGGCKRD 256  
Qy 117 LFPYSE--REANRSDGAIYVCYVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 167  
Db 257 PLEYSDVTPADAKA---GIIRYALGVDAFOEPTALKEINTIGSAPPDHFVKGN-FAA 312  
Qy 168 LOGIHSILKK 178  
Db 313 LRSIQRLQEK 323

RESULT 10  
US-09-350-259-37  
Sequence 37, Application US/09350259  
Patent No. 6620915  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 6620915e1 Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/350,259  
CURRENT FILING DATE: 1999-07-08  
EARLIER APPLICATION NUMBER: 09/193,043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 37  
LENGTH: 1151  
TYPE: PRT  
ORGANISM: Rattus rattus  
US-09-350-259-37

Query Match 14.6%; Score 141.5; DB 4; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;



Db 212 DPOSVDPIVQLQ-----GLTYTATGIRTWEELEFHSKNGSRKSAKKILLVITDGQKRD 266  
Qy 117 LFFYSB--REANSRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 167  
Db 267 PLEYSVDVIPADKX-----GIIRYALGVGDARFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 322  
Qy 168 LOGIHSILKK 178  
Db 323 LRSTQRQLQEK 333

## RESULT 13

US-08-605-672-55  
Sequence 55, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-55

Query Match 14.6%; Score 141.5; DB 2; Length 1161;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 4 DLYFIIDKSGSV-LAHMNEIYVVEQLAHKFI SPQLRMSFIVFSTGTTLMKLT-----57  
Db 152 DIAFLIDGSGSINQRDPAQWKDFV KALMGEPASTSTLFSIMQYSNLIKTHFTFEFNKIL 211  
Qy 58 DRQIRIGLELEQVLRGSDTYVHGEPRASQIYYENRGYRTA-SVIALITDGELEHD 116  
Db 212 DPOSVDPIVQLQ-----GLTYTATGIRTWEELEFHSKNGSRKSAKKILLVITDGQKRD 266

Qy 117 LFFYSB--REANSRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 167  
Db 267 PLEYSVDVIPADKX-----GIIRYALGVGDARFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 322  
Qy 168 LOGIHSILKK 178  
Db 323 LRSTQRQLQEK 333

## RESULT 14

US-08-482-293A-55  
Sequence 55, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-55

Query Match 14.6%; Score 141.5; DB 2; Length 1161;  
Best Local Similarity 27.7%; Pred. No. 4.5e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 4 DLYFIIDKSGSV-LAHMNEIYVVEQLAHKFI SPQLRMSFIVFSTGTTLMKLT-----57  
Db 152 DIAFLIDGSGSINQRDPAQWKDFV KALMGEPASTSTLFSIMQYSNLIKTHFTFEFNKIL 211  
Qy 58 DRQIRIGLELEQVLRGSDTYVHGEPRASQIYYENRGYRTA-SVIALITDGELEHD 116  
Db 212 DPOSVDPIVQLQ-----GLTYTATGIRTWEELEFHSKNGSRKSAKKILLVITDGQKRD 266  
Qy 117 LFFYSB--REANSRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 167

Db 267 PLEYSDVLPADKA--GIIRYAGVGDAFOEPTALKEINTIGSAPPDHFVKGN-FAA 322  
 QY 168 LOGIHSILKK 178  
 Db 323 LRSIQROLOEK 333

QY 168 LOGIHSILKK 178  
 Db 323 LRSIQROLOEK 333  
 Search completed: June 13, 2005, 20:05:09  
 Job time : 19.254 secs

RESULT 15  
 US-08-943-363-55  
 : Sequence 55, Application US/08943363  
 : Patent No. 5837478  
 : GENERAL INFORMATION:  
 : APPLICANT: Gallatin, W. Michael  
 : APPLICANT: Van der Vlieten, Monica  
 : TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit  
 : NUMBER OF SEQUENCES: 114  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 : STREET: 233 South Wacker Drive, 6300 Sear Tower  
 : CITY: Chicago  
 : STATE: Illinois  
 : COUNTRY: United States  
 : ZIP: 60606-6402  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentn Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/943,363  
 : FILING DATE:  
 : CLASSIFICATION: 530  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/173,497  
 : FILING DATE: 23-DEC-1993  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/286,889  
 : FILING DATE: 5-AUG-1994  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/362,652  
 : FILING DATE: 21-DEC-1994  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Williams Jr., Joseph A.  
 : REGISTRATION NUMBER: 38,659  
 : REFERENCE/DOCKET NUMBER: 27866/32684  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 312-474-6300  
 : TELEFAX: 312-474-0448  
 : TELEX: 25-3856  
 : INFORMATION FOR SEQ ID NO: 55:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1161 amino acids  
 : TYPE: amino acid  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-943-363-55

Query March 14.6%; Score 141.5; DB 2; Length 1161;  
 Best local similarity 27.7%; Pred. No. 4.5e-07;  
 Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
 QY 4 DLVFLDKSGSV-LHMNNEIYFVQOLAHKFIPLRMSFIYFSTRTGLMKLTE----- 57  
 Db 152 DIAFLIDSSGSLNQDFQMDPVKALMGEPASTSTLSLMQYSNILKTHFTPEFNIL 211  
 QY 58 DREQLRGLLELQKYLPGSDTYMHGFEFRASEQIYENRQGYRTA-SVIALTDGELHED 116  
 Db 212 DPOSILVPIVOLQ-----GLTYTATGIRTVMEELFHSKNGSRKSAKILVITDGQKRD 266  
 QY 117 LFFYSE--REANRSDIGAIYVCVGD-FNE-TQLARI-----ADSKDHFPVNDGFOA 167  
 Db 267 PLEYSDVLPADKA--GIIRYAGVGDAFOEPTALKEINTIGSAPPDHFVKGN-FAA 322



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 20:03:29 / Search time 59.9593 Seconds  
(without alignments)  
1195.537 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_41\_227

Perfect score: 970  
Sequence: 1 GGFPLYILDKSGSVLHMN.....LOGIHSILKSCIEILAE 187

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 383334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodaca/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodaca/1/pubppa/PTCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodaca/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodaca/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodaca/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodaca/1/pubppa/PTCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodaca/1/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodaca/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodaca/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodaca/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodaca/1/pubppa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodaca/1/pubppa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodaca/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodaca/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodaca/1/pubppa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodaca/1/pubppa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodaca/1/pubppa/US10\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodaca/1/pubppa/US10\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodaca/1/pubppa/US11\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodaca/1/pubppa/US11\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodaca/1/pubppa/US60\_NEW\_PUB.pep:\*  
22: /cgn2\_6/ptodaca/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	970	100.0	328	14	US-10-038-307-26
2	970	100.0	328	14	US-10-201-292-26
3	970	100.0	333	10	US-09-796-753-12
4	970	100.0	333	14	US-10-038-307-2
5	970	100.0	333	14	US-10-201-292-2
6	970	100.0	342	14	US-10-038-307-22
7	970	100.0	342	14	US-10-201-292-22
8	970	100.0	345	14	US-10-038-307-24
9	970	100.0	345	14	US-10-201-292-24
10	970	100.0	403	11	US-09-833-245-621
11	970	100.0	460	14	US-10-201-292-28

12	970	100.0	479	14	US-10-201-292-32	Sequence 32, App1
13	970	100.0	504	14	US-10-201-292-34	Sequence 34, App1
14	970	100.0	529	14	US-10-201-292-36	Sequence 36, App1
15	970	100.0	551	14	US-10-038-307-18	Sequence 18, App1
16	970	100.0	551	14	US-10-201-292-18	Sequence 18, App1
17	970	100.0	564	10	US-09-818-715-187	Sequence 187, App
18	970	100.0	564	10	US-09-818-715-232	Sequence 232, App
19	970	100.0	564	14	US-10-038-307-20	Sequence 20, App1
20	970	100.0	564	14	US-10-201-292-199	Sequence 199, App
21	970	100.0	564	14	US-10-038-307-12	Sequence 12, App
22	970	100.0	564	16	US-10-408-765-1823	Sequence 1823, App
23	970	100.0	564	16	US-10-474-794-187	Sequence 187, App
24	970	100.0	564	16	US-10-474-794-232	Sequence 232, App
25	965	99.5	403	11	US-09-833-245-620	Sequence 620, App
26	961	99.1	562	10	US-09-818-715-194	Sequence 194, App
27	961	99.1	562	10	US-09-818-715-301	Sequence 301, App
28	961	99.1	562	16	US-10-474-794-194	Sequence 194, App
29	961	99.1	562	16	US-10-474-794-301	Sequence 301, App
30	960	99.0	460	14	US-10-201-292-30	Sequence 30, App1
31	949	97.8	534	14	US-10-038-307-12	Sequence 12, App1
32	949	97.8	534	14	US-10-201-292-12	Sequence 12, App1
33	949	97.8	543	14	US-10-038-307-10	Sequence 10, App1
34	949	97.8	543	14	US-10-038-307-14	Sequence 14, App1
35	949	97.8	543	14	US-10-038-307-16	Sequence 16, App1
36	949	97.8	543	14	US-10-201-292-16	Sequence 16, App1
37	949	97.8	543	14	US-10-201-292-14	Sequence 14, App1
38	949	97.8	538	13	US-10-047-542-99	Sequence 99, App1
39	900	92.8	488	14	US-09-796-753-52	Sequence 52, App1
40	550	56.7	488	10	US-10-038-307-6	Sequence 6, App1
41	550	56.7	488	14	US-10-201-292-6	Sequence 6, App1
42	550	56.7	488	14	US-10-368-087-16	Sequence 16, App1
43	550	56.7	488	15	US-10-104-047-2639	Sequence 2639, App
44	550	56.7	488	15	US-09-796-753-54	Sequence 54, App1
45	548	56.5	487	10	US-09-796-753-54	Sequence 54, App1

## ALIGNMENTS

RESULT 1					
US-10-038-307-26					
Sequence 26, Application US/10038307					
Publication No. US20030134786A1					
GENERAL INFORMATION:					
APPLICANT: James B. ROTTMAN					
APPLICANT: Theresa L. O'KEEFE					
APPLICANT: Engin OZKAYNAK					
APPLICANT: Judith J. HEALEY					
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods					
FILE REFERENCE: 7853-253-999					
CURRENT APPLICATION NUMBER: US/10/038,307					
CURRENT FILING DATE: 2002-06-28					
NUMBER OF SEQ ID NOS: 26					
SOFTWARE: FASTSEQ for Windows Version 4.0					
SEQ ID NO 26					
LENGTH: 328					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-10-038-307-26					
Query Match					
Best Local Similarity 100.0%; Score 970; DB 14; Length 328;					
Pred. No. 6e-93;					
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY 1 GGFPLYILDKSGSVLHMNBEIYFVQLAHKPISPOLRMSFIYFSTRGTTMLKLTEDRE 60					
DB 41 GGFPLYILDKSGSVLHMNBEIYFVQLAHKPISPOLRMSFIYFSTRGTTMLKLTEDRE 100					
QY 61 QIRGAEELQKVPGGDTYHGEGERASEQIYYENRGYGTASYITLITGDELHEDLFY 120					
DB 101 QIRGAEELQKVPGGDTYHGEGERASEQIYYENRGYGTASYITLITGDELHEDLFY 160					
QY 121 SREANRSDLGAIVYCVGVDFNETOLARIADSKDHVPVNDGFOALOGIHSILKSC 180					

```
Db      161 SEREANSRDLGAIYVCVGVKDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSC 220
QY      181 IETIAAE 187
Db      221 IETIAAE 227

RESULT 2
US-10-201-292-26
; Sequence 26, Application US/10201292
; Publication No. US20030144193A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/201,292
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-292-26

Query Match      100.0%; Score 970; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 6e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GGFDFLYFIIDKSGSVLHNNIEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
Db      41 GGFDFLYFIIDKSGSVLHNNIEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100
QY      61 QIRQGLEELQKVLPGSDTYTMHGEFERASQIYYENRGYRTASVIALTDGELHEDLFFY 120
Db      101 QIRQGLEELQKVLPGSDTYTMHGEFERASQIYYENRGYRTASVIALTDGELHEDLFFY 160
QY      121 SEREANSRDLGAIYVCVGVKDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSC 180
Db      161 SEREANSRDLGAIYVCVGVKDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSC 220
QY      181 IETIAAE 187
Db      221 IETIAAE 227

RESULT 3
US-09-796-753-12
; Sequence 12, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
```

```
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 12
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-796-753-12

Query Match      100.0%; Score 970; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 6.2e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GGFDFLYFIIDKSGSVLHNNIEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
Db      41 GGFDFLYFIIDKSGSVLHNNIEIYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100
QY      61 QIRQGLEELQKVLPGSDTYTMHGEFERASQIYYENRGYRTASVIALTDGELHEDLFFY 120
Db      101 QIRQGLEELQKVLPGSDTYTMHGEFERASQIYYENRGYRTASVIALTDGELHEDLFFY 160
QY      121 SEREANSRDLGAIYVCVGVKDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSC 180
Db      161 SEREANSRDLGAIYVCVGVKDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSC 220
QY      181 IETIAAE 187
Db      221 IETIAAE 227

RESULT 4
US-10-038-307-2
; Sequence 2, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
```

```
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/038,307
/ CURRENT FILING DATE: 2002-06-28
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 333
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-038-307-2
```

```
Query Match      100.0%; Score 970; DB 14; Length 333;
Best Local Similarity 100.0%; Pred. No. 6,2e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60
DB 41 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100
QY 61 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 120
DB 101 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 160
QY 121 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 180
DB 161 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 220
QY 181 IEILAAE 187
DB 221 IEILAAE 227
```

```
RESULT 5
US-10-201-292-2
/ Sequence 22, Application US/10201292
/ Publication No. US20030144193A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/201,292
/ CURRENT FILING DATE: 2003-02-14
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 333
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-201-292-2
```

```
Query Match      100.0%; Score 970; DB 14; Length 333;
Best Local Similarity 100.0%; Pred. No. 6,2e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60
DB 41 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100
QY 61 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 120
DB 101 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 160
QY 121 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 180
DB 161 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 220
```

```
QY 181 IEILAAE 187
DB 221 IEILAAE 227
```

```
RESULT 6
US-10-038-307-22
/ Sequence 22, Application US/10038307
/ Publication No. US20030134786A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/038,307
/ CURRENT FILING DATE: 2002-06-28
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 342
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-038-307-22
```

```
Query Match      100.0%; Score 970; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 6,4e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60
DB 50 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 109
QY 61 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 120
DB 110 QIRQGLEELQKVLPGGDTYHHEGFERASEQIYYENRQGYRTASVITLTDGELHEDLFFY 169
QY 121 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 180
DB 170 SEREANSRDLGAIIVCVGKDFNETQTLARIADSKDHVPVNDGFQALOGIHSILKKSC 229
QY 181 IEILAAE 187
DB 230 IEILAAE 236
```

```
RESULT 7
US-10-201-292-22
/ Sequence 22, Application US/10201292
/ Publication No. US20030144193A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/201,292
/ CURRENT FILING DATE: 2003-02-14
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 342
/ TYPE: PRF
/ ORGANISM: Homo sapiens
US-10-201-292-22
```

```
Query Match      100.0%; Score 970; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 6,4e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GGFDFYFLDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60
```

```
Db 50 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 109
Qy 61 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 120
Db 110 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 169
Qy 121 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 180
Db 170 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 229
Qy 181 IETIAAE 187
Db 230 IETIAAE 226

RESULT 8
US-10-038-307-24
; Sequence 24, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 187 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038.307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-24

Query Match 100.0%; Score 970; DB 14; Length 345;
Best Local Similarity 100.0%; Pred. No. 6.5e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
Db 41 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100
Qy 61 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 120
Db 101 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 160
Qy 121 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 180
Db 161 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 220
Qy 181 IETIAAE 187
Db 221 IETIAAE 227

RESULT 9
US-10-201-292-24
; Sequence 24, Application US/10201292
; Publication No. US20030144193A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 187 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/201.292
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 24
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-292-24

Query Match 100.0%; Score 970; DB 14; Length 345;
Best Local Similarity 100.0%; Pred. No. 6.5e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
Db 41 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100
Qy 61 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 120
Db 101 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 160
Qy 121 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 180
Db 161 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 220
Qy 181 IETIAAE 187
Db 221 IETIAAE 227

RESULT 10
US-09-833-245-621
; Sequence 621, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833.245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 621
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-621

Query Match 100.0%; Score 970; DB 11; Length 403;
Best Local Similarity 100.0%; Pred. No. 8e-93;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
Db 41 GGFDFLYILDKSGSVLHHNMEIYFVBOIAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100
Qy 61 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 120
Db 101 QIRQGLEBLQKVLPGGDTYMHGEGFERASEQIYYENRQGYRTASVITALTGELHEDLFFY 160
Qy 121 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 180
Db 161 SEREANSRDLGAIYVCVGYKDFNETQIARIADSKDHVPVNDGFQALOGIISHILKSC 220
Qy 181 IETIAAE 187
Db 221 IETIAAE 227

RESULT 11
```

US-10-201-292-28  
; Sequence 28, Application US/10201292  
; Publication No. US20030144193A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/201,292  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 460  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-201-292-28

Query Match 100.0%; Score 970; DB 14; Length 460;  
Best Local Similarity 100.0%; Pred. No. 9,6e-93;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60  
DB 41 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100  
QY 61 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 120  
DB 101 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 180  
DB 161 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 220  
QY 181 IETILAE 187  
DB 221 IETILAE 227

RESULT 12  
US-10-201-292-32  
; Sequence 32, Application US/10201292  
; Publication No. US20030144193A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/201,292  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 479  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-201-292-32

Query Match 100.0%; Score 970; DB 14; Length 479;  
Best Local Similarity 100.0%; Pred. No. 1e-92;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60  
DB 41 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100  
QY 61 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 120  
DB 101 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 160

QY 121 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 180  
DB 161 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 220  
QY 181 IETILAE 187  
DB 221 IETILAE 227

RESULT 13  
US-10-201-292-34  
; Sequence 34, Application US/10201292  
; Publication No. US20030144193A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/201,292  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 504  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-201-292-34

Query Match 100.0%; Score 970; DB 14; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.1e-92;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 60  
DB 41 GGFDFYILDKSGSVLHNMNEIYFVEQLAHKFIISPOLRMSFIVFSTRGTTLMKLTEDRE 100  
QY 61 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 120  
DB 101 QIROGJLELOKVLPGGDTYHHEGFERASEQIYYENRQGYTASVITLTTGELHEDLFFY 160  
QY 121 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 180  
DB 161 SEREANSRDLGAIIVCVGVKDFNETQIARIADSKDHFVNDGFOALOGIHSILKXSC 220  
QY 181 IETILAE 187  
DB 221 IETILAE 227

RESULT 14  
US-10-201-292-36  
; Sequence 36, Application US/10201292  
; Publication No. US20030144193A1  
; GENERAL INFORMATION:  
; APPLICANT: James B. ROTTMAN  
; APPLICANT: Theresa L. O'KEEFE  
; APPLICANT: Engin OZKAYNAK  
; APPLICANT: Judith J. HEALEY  
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
; FILE REFERENCE: 7853-253-999  
; CURRENT APPLICATION NUMBER: US/10/201,292  
; CURRENT FILING DATE: 2003-02-14  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 36  
; LENGTH: 529  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-201-292-36

```
Query Match      100.0%; Score 970; DB 14; Length 529;
Best Local Similarity 100.0%; Pred. No. 1.2e-92;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GGFDLVFIIDKSGSVLHHMNEIYYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
DB      41 GGFDLVFIIDKSGSVLHHMNEIYYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100

QY      61 QIRQGLEELQKVLPGGDTYMHGEPERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 120
DB      101 QIRQGLEELQKVLPGGDTYMHGEPERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 160

QY      121 SEREANSRDLGAIIVCGVYKDFNETQLARIADSKDHVPVNDGFQALQGIHSHILKSC 180
DB      161 SEREANSRDLGAIIVCGVYKDFNETQLARIADSKDHVPVNDGFQALQGIHSHILKSC 220

QY      181 IEIILAE 187
DB      221 IEIILAE 227
```

```
RESULT 15
US-10-038-307-18
; Sequence 18; Application US/10038307
; Publication No US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Englin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038.307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-18
```

```
Query Match      100.0%; Score 970; DB 14; Length 551;
Best Local Similarity 100.0%; Pred. No. 1.2e-92;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 GGFDLVFIIDKSGSVLHHMNEIYYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 60
DB      41 GGFDLVFIIDKSGSVLHHMNEIYYFVEQLAHKFISPOLRMSFIVFSTRGTTLMKLTEDRE 100

QY      61 QIRQGLEELQKVLPGGDTYMHGEPERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 120
DB      101 QIRQGLEELQKVLPGGDTYMHGEPERASEQIYYENRQGYRTASVIALTDGELHEDLFFY 160

QY      121 SEREANSRDLGAIIVCGVYKDFNETQLARIADSKDHVPVNDGFQALQGIHSHILKSC 180
DB      161 SEREANSRDLGAIIVCGVYKDFNETQLARIADSKDHVPVNDGFQALQGIHSHILKSC 220

QY      181 IEIILAE 187
DB      221 IEIILAE 227
```

Search completed: June 13, 2005, 20:36:50  
Job time : 59.9593 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: June 13, 2005, 19:47:08 ; Search time 13.4163 Seconds  
(without alignments)  
1341.095 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_41\_227  
Perfect score: 970  
Sequence: 1 GGFGLYFLDKSGSVLHMN.....LOGIHSILKSCIEILAE 187

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	14.1	724	2 A48569	antigen Em100 - E1
2	132.5	13.7	1153	2 A48569	cell surface glyco
3	129	13.3	3124	2 A40020	collagen alpha 1(X
4	128.5	13.2	712	2 A45638	immunoglobulin m
5	125.5	12.9	1163	1 RWHU1C	cell surface glyco
6	124.5	12.8	1153	2 S00551	leukocyte surface
7	118.5	12.2	760	1 C2MS	classical-comple
8	116	12.0	3051	2 S42373	hypothetical prote
9	113	11.6	1747	2 A45974	collagen alpha 1(X
10	113	11.6	1857	2 S31212	collagen alpha 1(X
11	113	11.6	1888	2 S78476	collagen alpha 1(X
12	111.5	11.5	764	1 BBHU	complement factor
13	111.5	11.5	929	2 A54849	type XII collagen
14	110	11.3	2944	2 A54849	collagen alpha 1(V
15	109.5	11.3	1029	1 S21369	collagen alpha 2(V
16	103.5	10.7	567	2 T28797	hypothetical prote
17	103.5	10.7	761	1 BBMS	hypothetical prote
18	100	10.3	460	2 T23087	hypothetical prote
19	99	10.2	698	2 B96958	dnak protein (heat
20	99	10.2	917	2 S09646	collagen alpha 2(V
21	99	10.2	1018	1 CGH2A	collagen alpha 2(V
22	98	10.1	741	2 T46488	hypothetical prote
23	97.5	10.1	496	2 A37979	cartilage matrix p
24	96.5	9.9	493	2 A33809	cartilage matrix p
25	96.5	9.9	574	2 A46283	sporozoite surface
26	95.5	9.8	500	2 S66522	cartilage matrix p
27	95.5	9.8	559	2 S04531	thrombospondin-rel
28	95.5	9.8	1179	2 A53213	integrin alpha-B-c
29	94	9.7	3176	2 CGH2A	collagen alpha 3(V

30	93.5	9.6	843	2 A40970	undulin 1 - human
31	93	9.6	340	2 E70121	hypothetical prote
32	92.5	9.5	3137	2 A37797	collagen alpha 3(V
33	88.5	9.1	932	2 JC5953	inter-alpha-inhibi
34	88	9.1	1218	2 T30889	valine-tRNA ligase
35	87.5	9.0	341	2 T32949	hypothetical prote
36	87.5	9.0	418	2 AG2350	hypothetical prote
37	87	9.0	710	2 AG2285	hypothetical prote
38	85.5	8.8	340	2 B83285	hypothetical prote
39	85.5	8.8	380	2 H90261	hypothetical prote
40	85	8.8	1021	2 T05901	hypothetical prote
41	84.5	8.7	689	2 F84811	probable retroelem
42	84.5	8.7	800	2 AB1079	endoglucanase homo
43	84.5	8.7	875	1 ITECAP	DNA topoisomerase
44	84.5	8.7	875	2 B91018	DNA gyrase subunit
45	84.5	8.7	875	2 D85862	DNA gyrase, subunit

## ALIGNMENTS

### RESULT 1

A48569  
antigen Em100 - Eimeria maxima  
C/Species: Eimeria maxima  
C/Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A48569  
R:Pasamonies, L., Hug, D., Humbelin, M., Weber, G.  
Mol. Biochem. Parasitol. 57, 171-174, 1993  
A/Title: Sequence of a major Eimeria maxima antigen homologous to the Eimeria tenella m  
A/Reference number: A48569; MUID:93149203; PMID:8426611  
A/Accession: A48569  
A/Status: preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1-724 <PMS>  
A/Cross-References: UNIPROT: Q04588; GB: M99058; NID: G158890; PTD: G158891  
A/Note: Sequence extracted from NCBI backbone (NCBI:123776, NCBI:123777)  
F/45-218/Domain: von Willebrand factor type A repeat homology <VWA2>  
F/238-296/Domain: thrombospondin type 1 repeat homology <THR1>  
F/309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
F/372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
F/433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
F/494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
F/560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 14.1%; Score 137; DB 2; Length 724;  
Best Local Similarity 26.6%; Pred. No. 0.0007;

Matches 53; Conservative 38; Mismatches 80; Indels 28; Gaps 12;

QY	4	DLYFLDKSGSV-LHNMNEIYFVEQLAHKF-ISP-QLRMSFVSTRTLMKLTSDRE	60
DB	47	DWMLVDESSIGTSNGKRSFISNFGTMRPLSPDVRGVLTFGSAVTRMDSRA	106
QY	61	QIRQGLEELQKLP-P-GGDTYMEGFERASEQIYENRQGR--TASVIALTDGLHED	116
DB	107	QNMDLAAAKKLPYAAAGSTYTHLGLAKA-EELIFSGKGRDVAPEMLVTMDGA----	161
QY	117	LFYFSERE-----ANRSRLGAIYVCVQ-KQFNELQARLA--DKDNV-PE--VNDG	164
DB	162	-----SSRSQTLNAEKLRNRGVIIVLVGVGTGVSNAECSRIAGCDTSYVECPYLOSN	217
QY	165	FOALOGIHSILKSCIEI 183	
DB	218	WGVSSQINIGIRAKCDL 236	

### RESULT 2

RWHU1B  
cell surface glycoprotein CD11b precursor (validated) - human  
N/Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac  
eukocyte integrin alpha chain; neutrophil adherence receptor alpha chain  
C/Species: Homo sapiens (man)  
C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004

C/Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; I52567  
 R/Corby, A.L.; Kishimoto, T.K.; Miller, U.J.; Springer, T.A.  
 J. Biol. Chem. 263, 12403-12411, 1988  
 A/Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11B)  
 A/Reference number: A31108; MUID:88315033; PMID:2457584  
 A/Accession: A31108  
 A/Molecule type: mRNA  
 A/Residues: 1-1153 <COR>  
 A/Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148  
 A/Note: part of this sequence was confirmed by protein sequencing  
 R/Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
 J. Cell Biol. 106, 2153-2158, 1988  
 A/Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M  
 A/Reference number: A28915; MUID:88257215; PMID:2454931  
 A/Accession: A28915  
 A/Molecule type: mRNA  
 A/Residues: 1-499,501-965,'P',967-1153 <ARN>  
 A/Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594  
 A/Note: the authors translated the codon TAC for residue 1129 as Thr  
 R/Shelley, C.S.; Arnaout, M.A.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
 A/Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
 A/Reference number: A41600; MUID:92073318; PMID:1683702  
 A/Accession: A41600  
 A/Molecule type: DNA  
 A/Residues: 1-9 <SHE>  
 A/Cross-references: GB:M76724; NID:G180018; PIDN:AAA58410.1; PID:G553215  
 R/Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
 A/Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion  
 A/Reference number: A94193; MUID:86190151; PMID:2853753  
 A/Accession: A30892  
 A/Molecule type: mRNA  
 A/Residues: 917-1042 <AR2>  
 A/Cross-references: GB:M18044  
 R/Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
 A/Title: CDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
 A/Reference number: A32218; MUID:89098893; PMID:2563162  
 A/Accession: A32218  
 A/Molecule type: mRNA  
 A/Residues: 9-1153 <HIC>  
 A/Cross-references: GB:U04145; NID:G189068; PIDN:AAA59903.1; PID:G386975  
 A/Note: part of this sequence was confirmed by protein sequencing  
 R/Pleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
 J. Immunol. 150, 480-490, 1993  
 A/Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1  
 n during evolution.  
 A/Reference number: A46526; MUID:93123748; PMID:8419480  
 A/Accession: A46526  
 A/Molecule type: DNA  
 A/Residues: 1-499,501-1153 <PLE>  
 A/Cross-references: GB:S52227; NID:G263047; PIDN:ABB24821.1; PID:G263049  
 A/Note: the last three bases of Intron 13, CAG, are included in some but not all mature  
 A/Note: sequence extracted from NCBI backbone (NCBI:P.121963)  
 R/Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.  
 Biochim. Biophys. Acta 874, 368-371, 1986  
 A/Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
 A/Reference number: A90664; MUID:87076671; PMID:3539202  
 A/Accession: A90664  
 A/Molecule type: protein  
 A/Residues: 17-31 <PIE>  
 A/Molecule type: protein  
 A/Experimental source: granulocytes  
 R/Pahl, H.L.; Rosemarin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992  
 A/Title: Characterization of the myeloid-specific CD11b promoter.  
 A/Reference number: I52567; MUID:92144986; PMID:11346576  
 A/Accession: I52567  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA

A/Residues: 1-9 <RES>  
 A/Cross-references: GB:M4477; NID:G180184; PIDN:AAA51960.1; PID:G553219  
 C/Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1 c  
 C/Genetic:  
 A/Gene: GDB:ITGAM, CR3A  
 A/Cross-references: GDB:120599; OMIM:120980  
 A/Map position: 16p11.2-16p11.2  
 A/Note: promoter contains a GATA motif and two Sp1 consensus binding sites  
 C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homol  
 C/Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; magr  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>  
 F:17-1108/Domain: extracellular #status predicted <EXT>  
 F:148-318/Domain: von Willebrand factor type A repeat homology <VMA2>  
 F:465-473/Region: calcium/magnesium binding #status predicted  
 F:530-538/Region: calcium/magnesium binding #status predicted  
 F:593-601/Region: calcium/magnesium binding #status predicted  
 F:1109-1134/Domain: transmembrane #status predicted <TM>  
 F:1135-1153/Domain: intracellular #status predicted <INT>  
 F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
 Query Match 13.7%; Score 132.5; DB 1; Length 1153;  
 Best local similarity 26.3%; Pred. No. 0.003;  
 Matches 52; Conservative 41; Mismatches 66; Indels 39; Gaps 11;  
 QY 4 DLYFIIDKGSVV-HHNNLYVVEQLAKFISPOLRMSTIVSTGTTMKLTED---- 58  
 DB 150 DIAFLIDSGSIIIPDFRKRKEFVST---VNEOLKSKSTLPS---LMOYSEPRRIH 199  
 QY 59 -----RQIQGLEELQKVPGSDTYVHEGFERSAQIYENRGVYTA-SVITALT 109  
 DB 200 FTFKEQNNPNPNSLVKPIITQL--GRTHATYIRKVRRELPIITGARGAKKILIVIT 257  
 QY 110 DGEIHEDLFFYS--RANKSRDLGAIYVCVGFENETQLAR----IADS--KDHVP 160  
 DB 258 DGEKFDGPEYEDVIDEADRE---GVIRYVIGVGDAPFRSEKSHQELNTIASKPPRDHVFQ 314  
 QY 161 VNDGFQALGCIHTSIILK 178  
 DB 315 VNN-FRALKTIOMQLREK 331  
 RESULT 3  
 A40020  
 collagen alpha 1(XII) chain precursor - chicken  
 N/Alternate names: fibrochmerin  
 C/Species: Gallus gallus (chicken)  
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004  
 C/Accession: A40020; A34485; B34485; A28037; S23814; S22554; S28811  
 R/Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obari  
 J. Cell Biol. 115, 209-221, 1991  
 A/Title: The complete primary structure of type XII collagen shows a chimeric molecule w  
 nou region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.  
 A/Reference number: A40020; MUID:92011862; PMID:1918137  
 A/Accession: A40020  
 A/Molecule type: mRNA  
 A/Residues: 1-3124 <YAM>  
 A/Cross-references: UNIPROT:P13944; GB:D00824; NID:G222810; PIDN:BA00701.1; PID:G222811  
 A/Note: in the authors' translation residues 1216-1219 are shown after residue 1235 and,  
 R/Gordon, M.K.; Gerecke, D.R.; Dabler, B.; van der Rest, M.; Olsen, B.R.  
 J. Biol. Chem. 264, 19772-19778, 1989  
 A/Title: Type XII collagen. A large multidomain molecule with partial homology to type I  
 A/Reference number: A34485; MUID:9062079; PMID:2584192  
 A/Accession: A34485  
 A/Molecule type: mRNA  
 A/Residues: 2456-2758 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>  
 A/Cross-references: EMBL:U05137; NID:G211284; PIDN:AAA46835.1; PID:G211285  
 A/Accession: B34485  
 A/Molecule type: protein  
 A/Residues: 2772-2792; 2846-2873 <GOR2>  
 R/Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987  
 A/Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA cl





A/Reference number: A36584  
A/Contents: erratum  
A/Accession: A36584  
A/Molecule type: DNA  
A/Residues: 1-1163 <COR>  
A/Cross-references: UNIPROT:P20702  
A/Note: this revision to the sequence from reference A35543 includes the carboxyl end  
R/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2788, 1990  
A/Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A/Reference number: A35543; MUID:90153906; PMID:2303426  
A/Accession: A35543  
A/Molecule type: DNA  
A/Residues: 1-834 <CO2>  
A/Note: this sequence has been revised in reference A36584  
R/Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
EMBO J. 6, 4033-4028, 1987  
A/Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
A/Reference number: 500664; MUID:88166645; PMID:3327687  
A/Accession: 500664  
A/Molecule type: mRNA  
A/Residues: 1-755; 'L', 757-1163 <CO3>  
A/Cross-references: GB:M81695; EMBL:Y00093; NID:G487829; PIDN:AAA59180.1; PID:G487830  
A/Note: part of this sequence was confirmed by protein sequencing  
C/Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my  
C/Genetics: GDB:ITGAX; CD11C  
A/Gene: GDB:ITGAX; CD11C  
A/Cross-references: GDB:119758; OMIM:151510  
A/Map position: 16p11.2-16p11.2  
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
C/Keywords: cell adhesion; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>  
F:149-1107/Domain: extracellular #status predicted <EXT>  
F:1108-1133/Domain: von Willebrand factor type A repeat homology <VMA4>  
F:1134-1163/Domain: transmembrane #status predicted <TM>  
F:61,89,392,657,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre  
Query Match 12.9%; Score 125.5; DB 1; Length 1163;  
Best Local Similarity 24.7%; Pred. No. 0.012;  
Matches 48; Conservative 38; Mismatches 77; Indels 31; Gaps 9;  
QY 4 DLYFILDKSGSV-LHMHNEIYFVQLAHKFTISQLMSFVFSRTGTTMLKLTEDR-- 60  
DB 151 DIVFLIDGSGSISRNPAWNNFRAVISQFQRPSTQFSLMQFSNKEQTFTEPEEFRTS 210  
QY 61 ---QIRQGLEBLQKVLPGSDTYMHGEFPASEQIYENRQGYRAS-VIITALDGEIHED 116  
DB 211 NPISLASVHQLQ-----GFTYTAIAQNVVHRLFHASVYGARRDATKILVITDGKKEGD 265  
QY 117 LFYSEREANRSRDGAIVYCVG-----KDFNETQLARIAD--SKDHVPVNDG 164  
DB 266 SLDYKD-VIMADAGIIRAIAGVGLAFQNRMSKELND-----IAKPSQEHIFKED- 318  
QY 165 FQALQGIHHSILKK 178  
DB 319 FDLAKDIONQLKEX 332  
RESULT 6  
S00551  
Leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse  
N/Alternate names: complement-3 receptor alpha chain  
C/Species: Mus musculus (house mouse)  
C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C/Accession: S00551; I55078  
R/Pytela, R.  
EMBO J. 7, 1371-1378, 1988  
A/Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the  
A/Reference number: S00551; MUID:88312584; PMID:3044779  
A/Accession: S00551  
A/Molecule type: DNA

A/Residues: 1-1153 <PYT>  
A/Cross-references: UNIPROT:P05555; EMBL:X07640; NID:9552982; PIDN:CAA30479.1; PID:g552983  
A/Note: the authors translated the codon CAC for residue 569 as Gln  
R/Sastre, L.; Roman, J.M.; Teplov, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,  
Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986  
A/Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recept  
A/Reference number: I59078; MUID:86287312; PMID:2942940  
A/Accession: I59078  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 11-44 <RES>  
A/Cross-references: GB:M14293; NID:g198993; PIDN:AAA39484.1; PID:g554193  
C/Genetics:  
A/Gene: Mac-1  
C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
C/Keywords: cell adhesion; glycoprotein; transmembrane protein  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental  
F:148-118/Domain: von Willebrand factor type A repeat homology <VMA2>  
F:1106-1129/Domain: transmembrane #status predicted <TM>  
Query Match 12.8%; Score 124.5; DB 2; Length 1153;  
Best Local Similarity 26.8%; Pred. No. 0.014;  
Matches 53; Conservative 37; Mismatches 69; Indels 39; Gaps 12;  
QY 4 DLYFILDKSGSV-LHMHNEIYFVQLAHKFTIS--PQLRMSFVFSRTGTTMLKLTED-- 58  
DB 150 DIVFLIDGSGSI-----NNIDF--QKMKFVSTVMEQFKSKTLFS-----LMQYSDFFR 197  
QY 59 ---REQIRGLEBLQKVLPGSDTYMHGEFPASEQIYENRQGYRAS-VIITALD 109  
DB 198 IHTFNDPKRNPSPRSHSVSYKQNGRTKTSAGIRKYRELFFKNGARENAKILVIT 257  
QY 110 DGEIHEDLFYYS--REANRSRDGAIVYCVGYKD--FNETQLARIAD-----SKDHVP 160  
DB 258 DGEKFGDPLDYKQVTEADRA---GIRYIVIGVGNAPNKRQSRBELDTIASKPAGENVFQ 314  
QY 161 VNDGFQALQGIHHSILKK 178  
DB 315 V-DNFEALNTIONQLQEK 331  
RESULT 7  
C2N5  
Classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2 component precursor - mor  
N/Alternate names: C3 convertase; C5 convertase; complement C2  
C/Species: Mus musculus (house mouse)  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: A38876; B36593; I54429  
R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
submitted to Genbank, January 1991  
A/Reference number: A38875  
A/Accession: A38875  
A/Molecule type: DNA  
A/Residues: 1-760 <IS2>  
A/Cross-references: UNIPROT:P21180; GB:M57891; GB:J05661; NID:g192436; PIDN:AAA63294.1; I  
R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
J. Biol. Chem. 265, 19040-19046, 1990  
A/Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different me  
A/Reference number: A36593; MUID:91035430; PMID:2229060  
A/Accession: B36593  
A/Molecule type: mRNA  
A/Residues: 1-760 <ISH>  
A/Cross-references: EMBL:M57891; NID:g192436; PIDN:AAA63294.1; PID:g192437  
R/Falusi, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, J.; Whitehead, A.S.; Colten, H.R.  
Immunogenetics 25, 290-298, 1987  
A/Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.  
A/Reference number: I54429; MUID:87192938; PMID:2883115  
A/Accession: I54429  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 660-677; 'R', 679, 681-723; 'G', 725 <RES>  
A/Cross-references: GB:M16271; NID:g199289; PIDN:AAA9562.1; PID:g199290

C/Genetics: A:Interons: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 414/1; 461/1; 492/3; 554/1; 570/689/Active site: His, Asp, Ser #status predicted  
C/Complex: The proenzyme forms a complex with C2a and is activated by cleavage into C2a.  
C/Function: A:Description: cleaves complement C3 and complement C5 alpha chains  
A:Pathway: complement classical pathway  
C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; \n  
C/Keywords: alternative splicing; complement C2b fragment #status predicted <SIG>  
F/1-18/Domain: signal sequence #status predicted <C2B>  
F/19-250/Product: complement C2b fragment #status predicted <C2B>  
F/19-250/Product: complement C2b fragment #status predicted <C2B>  
F/22-89/Domain: complement factor H repeat homology <FH1>  
F/94-149/Domain: complement factor H repeat homology <FH2>  
F/156-210/Domain: complement factor H repeat homology <FH3>  
F/251-760/Product: complement C2a fragment long form #status predicted <C2a>  
F/251-605,613-760/Product: complement C2a fragment short form #status predicted <C2s>  
F/259-449/Domain: von Willebrand factor type A repeat homology <VFA>  
F/428-747/Domain: trypsin homology #status atypical <TRY>  
F/22-62,44-89, 94-136, 122-149, 156-197, 182-210, 470-590, 499-515, 593-609, 647-674, 685-715/DiDisulfide bond: 12/1; 297, 340-474, 478, 669/Binding site: carbohydrate (am) (covalent) #status predicted <C2a>  
F/514,570,689/Active site: His, Asp, Ser #status predicted

		Query Match	Similarity	12.2%	Score 118.5	DB 1	length 760;
		Best Local	Similarity	24.4%	Pred. No. 0.029		
		Matches	53	Conservative	41	Mismatches	74; Indels 49; Gaps 10
Qy	1	GGFDDYFLIDSGSVLHMNRIY-YFVEQLAHKFPISPOLR-MSFIVFSTRTGLTKLTE	57				
Db	258	GHLMNYLLDLSQSOTEKDFIKFSABEMVERIFSFENVVVAITTFASQPPTIMSILS	317				
Qy	58	DREOIRGQLEBELQKLPGCDTYMHGEFEASEIYYE-----NRGYNITAS-	103				
Db	318	ERSQC-----DTTEVTISLDSSAYKDHEATANTAEVLIRVSYMMOTOMDRIGMETSAW	371				
Qy	104	-----VIALTDEGLEHEDLFYSERERANSRDIGAI-----YCUGVV-----XD	142				
Db	372	KEIHRTIIILLTGK--SNMGDSPKKAVTRIRELTSLIEQRNDVDLYALGVGLDVDMWKE	429				
Qy	143	FNEQTOLARIADSKDHVPVNDSFGPQAQGIIHSILKKS	179				
Db	430	LNE--LGSKKDGGERHAFFILODA-KALLQQCFEHHLDVDVS	463				

```

RESULT 8
S42373
hypothetical protein T20G5.3 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 09-May-2004
C|Accession: S42373
R|Smith, A.
submitted to the EMBL Data Library, March 1994
A|Reference number: S42373
A|Accession: S42373
A|Molecule type: DNA
A|Residues: 1-3051 <SWI>
A|Cross-references: EMBL:Z30423; NID:G458479; PID:G458485
C|Genetic8:
A|Intercom8: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/1; 952/2; 1041/1; 1146/1; 1512-6/9/Domains: von Willebrand factor type A repeat homology <WFA1>
F|754-793/Domains: fibronectin type II repeat homology <2F1>
F|1201-1244/Domains: EGF homology <EGF>

```

[illegible]

**Oy** 106 IATDELHEDJFEYSEARSRLGALVCGVKD-FNSTOLARLIDSQKHFPVNDG 164  
| | | | : : : :  
**Db** 624 ILTDERSDQNV-----TGPSDARSLSTNTPALIGVTIDVLASLELSIAGSNNRPFVY-DK 678  
| | | | : : : :  
**Oy** 165 FOALOGIIHSILKLS 179  
| : : : : : : : :  
**Db** 679 FKDLDTLRKSMIOKA 693

```

RESULT 9
A45974
collagen alpha 1(XIV) chain precursor, short form 2 - chicken
N:Alternate names: undulin
C:Species: Gallus gallus (chicken)
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A45974; S30085; S22916; S17035; S20833
R:Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Lin, J. Biol. Chem. 268, 12177-12184, 1993
A:Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' regions

```

A:Accession number: A45974; MUID:93280195; PMID:850033/  
A:Accession: A45974  
A:Status: preliminary  
A:Molecule type: mRNA; protein  
A:Residues: 1-1747 <CGR>  
A:Cross-references: UNIPROT:P32018  
A:Experimental source: embryo skin  
A>Note: sequence inconsistent with the nucleotide translation  
A>Note: sequence extracted from NCBI Backbone (NCBIN:13364, NCBIP:13365  
R:Ape, S.S.  
submitted to the EMBL Data Library, March 1992  
A:Reference number: S30085  
A:Accession: S30085  
A:Molecule type: mRNA  
A:Residues: 1472-1660 <APT>  
A:Cross-references: EMBL:X65122; NID:G62871; PIDN:CAA46238.1; PID:G938175  
R:Trueb, J., Trueb, B.  
Eur. J. Biochem. 207, 549-557, 1992  
A:Title: Type XIV collagen is a variant of undulin.  
A:Reference number: S22916; MUID:92339443; PMID:1339349

Query Match	11.6%;	Score 113;	DB 2;	Length 1747;
Best Local Similarity	26.1%;	Pred. No. 0.23;		
Matches	49;	Conservative	39;	Mismatches
			80;	Indels
			20;	Gaps
			11;	

4 DLYFIIDKSSGV-LHHNNEIYFVEQL--AKRFSPO-LRMSFVFSRGTLLMKLT--E 57

```
Db      926 DLVFLVGSWSIGDNNFKIISFLYSTVGALDKIGPGCTOVAIIQFSDDPRTTEKLNAYK 985
      58 DREGIROGLEBELQKVLPGDITVMEHGFERSAQIYYENRQGYRTA--SVIIALTDSGLH 115
      986 TKETLLEAIQOI--AYKGGNTKTGKAIKHARE-VLFTGEGAMRGKIGPKVLVITDGRSQD 1042
Qy      116 DLFPYSERANRSDI-GAIVYCVGVDFNETOLARIAD--SKDHVPVNDGFOALOGII 172
      1043 DV-----NKYSREMOIDGFSFPAIGVADADYSELVNIWGSKPSRHHVFVVD-  
FDPATFKIE 1096
Qy      173 HSLIKKSC 180
      1097 DELITPVC 1104
```

## RESULT 10

```
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 15-Sep-2003
C/Accession: S31212
R/Melchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185666; PMID:8444186
A/Accession: S31212
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 1-1857 <WAE>
A/Cross-references: EMBL:X70792; NID:9288874; PIDN:CAA50064.1; PID:9288875
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C/Genetics:
A/Gene: Col14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F/1-28/Domain: signal sequence #status predicted <SIG>
F/29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F/29-110/Domain: fibronectin type III repeat homology <FN3A>
F/156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F/352-433/Domain: fibronectin type III repeat homology <FN3B>
F/442-525/Domain: fibronectin type III repeat homology <FN3C>
F/534-514/Domain: fibronectin type III repeat homology <FN3D>
F/623-707/Domain: fibronectin type III repeat homology <FN3E>
F/741-823/Domain: fibronectin type III repeat homology <FN3F>
F/832-914/Domain: fibronectin type III repeat homology <FN3G>
F/922-1009/Domain: fibronectin type III repeat homology <FN3H>
F/1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>
```

Query Match 11.6%; Score 113; DB 2; Length 1857;

Best Local Similarity 26.1%; Pred. No. 0.25; Matches 49; Conservative 39; Mismatches 80; Indels 20; Gaps 11;

```
Qy      4 DLVFIIDKSGSV-LHHNNEIYFVEOL--AHKFIQ-LRMSFIVFSTRTGTTMLKLT--E 57
      1042 DLVFLVGSWSIGDNNFKIISFLYSTVGALDKIGPGCTOVAIIQFSDDPRTTEKLNAYK 1101
Db      58 DREGIROGLEBELQKVLPGDITVMEHGFERSAQIYYENRQGYRTA--SVIIALTDSGLH 115
      1102 TKETLLEAIQOI--AYKGGNTKTGKAIKHARE-VLFTGEGAMRGKIGPKVLVITDGRSQD 1158
Qy      116 DLFPYSERANRSDI-GAIVYCVGVDFNETOLARIAD--SKDHVPVNDGFOALOGII 172
      1159 DV-----NKYSREMOIDGFSFPAIGVADADYSELVNIWGSKPSRHHVFVVD-  
FDPATFKIE 1212
Qy      173 HSLIKKSC 180
      1213 DELITPVC 1220
```

RESULT 11  
S78476  
collagen alpha 1(XIV) chain precursor, long form - chicken  
C/Species: Gallus gallus (chicken)

C/Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 09-Jul-2004  
C/Accession: S78476; S31211  
R/Trueb, B.

submitted to the EMBL Data Library, January 1993  
A/Reference number: S78476  
A/Accession: S78476

A/Molecule type: mRNA  
A/Residues: 1-1888 <TRU>

A/Cross-references: UNIPROT:P32018; EMBL:X70793; NID:9288872; PIDN:CAA50064.1; PID:9288875  
R/Melchli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.

Eur. J. Biochem. 212, 483-490, 1993  
A/Title: Complete primary structure of chicken collagen XIV.

A/Reference number: S31211; MUID:93185666; PMID:8444186  
A/Accession: S31211

A/Status: preliminary  
A/Molecule type: mRNA

A/Residues: 1-416;1460-1811,1843-1888 <WAE>  
A/Cross-references: EMBL:X70793

C/Genetics:

A/Gene: Col14A1

C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F/1-28/Domain: signal sequence #status predicted <SIG>

F/29-1888/Product: collagen alpha 1(XIV) chain, long form #status predicted <MAT>  
F/29-110/Domain: fibronectin type III repeat homology <FN3A>

F/156-320/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/352-433/Domain: fibronectin type III repeat homology <FN3B>

F/442-525/Domain: fibronectin type III repeat homology <FN3C>  
F/534-514/Domain: fibronectin type III repeat homology <FN3D>

F/623-707/Domain: fibronectin type III repeat homology <FN3E>  
F/741-823/Domain: fibronectin type III repeat homology <FN3F>

F/832-914/Domain: fibronectin type III repeat homology <FN3G>  
F/922-1009/Domain: fibronectin type III repeat homology <FN3H>

F/1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 11.6%; Score 113; DB 2; Length 1888;

Best Local Similarity 26.1%; Pred. No. 0.25; Matches 49; Conservative 39; Mismatches 80; Indels 20; Gaps 11;

```
Qy      4 DLVFIIDKSGSV-LHHNNEIYFVEOL--AHKFIQ-LRMSFIVFSTRTGTTMLKLT--E 57
      1042 DLVFLVGSWSIGDNNFKIISFLYSTVGALDKIGPGCTOVAIIQFSDDPRTTEKLNAYK 1101
Db      58 DREGIROGLEBELQKVLPGDITVMEHGFERSAQIYYENRQGYRTA--SVIIALTDSGLH 115
      1102 TKETLLEAIQOI--AYKGGNTKTGKAIKHARE-VLFTGEGAMRGKIGPKVLVITDGRSQD 1158
Qy      116 DLFPYSERANRSDI-GAIVYCVGVDFNETOLARIAD--SKDHVPVNDGFOALOGII 172
      1159 DV-----NKYSREMOIDGFSFPAIGVADADYSELVNIWGSKPSRHHVFVVD-  
FDPATFKIE 1212
Qy      173 HSLIKKSC 180
      1213 DELITPVC 1220
```

## RESULT 12

BBHU  
complement factor B precursor [validated] - human

N/Alternate names: C3 convertase; C3 proactivator; glycoline-rich beta-glycoprotein; heat-labile complement alternative-complement-pathway C3/5 convertase (EC 3.4.21.47) Bb fragment

C/Species: Homo sapiens (hmn)  
C/Date: 19-Feb-1984 #sequence\_revision 05-Aug-1994 #text\_change 09-Jul-2004

C/Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628; I5444  
R/Mejia, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.

submitted to the EMBL Data Library, March 1993  
A/Reference number: S34075

A/Accession: S34075  
A/Molecule type: mRNA

A/Residues: 1-764 <MEJ>  
A/Cross-references: UNIPROT:P00751; EMBL:X72875; NID:9297568; PIDN:CAA51389.1; PID:9297568  
R/Moore, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.

Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
A/Title: Isolation of cDNA clones for the human complement protein factor B, a class III

A/Reference number: A44622; MUID:83039428; PMID:6957884  
A/Accession: A44622  
A/Molecule type: mRNA  
A/Residues: 467-546/550-595/752-764 <MO>  
A/Cross-references: GB:U00185; GB:U00186  
R/Note: The authors translated the codon TAC at 519 as Thr: the nucleic acid translation  
U. Biol. Chem. 259, 3407-3412, 1984  
A/Title: Complete primary structure for the zymogen of human complement factor B.  
A/Reference number: A20751; MUID:84161997; PMID:6546754  
A/Accession: A00934  
A/Molecule type: Protein; mRNA  
A/Residues: 26-764 <MO>  
A/Cross-references: GB:K01566  
A/Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
A/Note: glycosylation sites were determined  
R/Chem. J. 209, 61-70, 1983  
A/Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th  
A/Reference number: A19188; MUID:83204002; PMID:6342610  
A/Contents: the final paper in a series documenting the sequence, glycosylation site, an  
A/Molecule type: Protein  
A/Residues: 260-286/298-764 <CHR>  
R/Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A/Title: Molecular cloning and characterization of the gene coding for human complement  
A/Reference number: A19947; MUID:83273641; PMID:6308626  
A/Accession: A19947  
A/Molecule type: DNA  
A/Residues: 346-764 <CAM>  
A/Cross-references: GB:U00125  
A/Accession: B19947  
A/Molecule type: mRNA  
A/Residues: 339-509 <CA1>  
A/Cross-references: GB:U00126; NID:G187723; PIDN:AAA36226.1; PID:G553536  
R/Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A/Title: Cell-specific expression of the human complement protein factor B gene: evidence  
A/Reference number: A25971; MUID:87102880; PMID:3643061  
A/Accession: B25971  
A/Molecule type: DNA  
A/Residues: 1-99 <MD>  
A/Cross-references: GB:M15082; NID:G187699; PIDN:AAA59625.1; PID:G553534  
R/Niemann, M.A.; Bhown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A/Title: The principal site of glycation of human complement factor B.  
A/Reference number: S14339; MUID:91174758; PMID:2006911  
A/Accession: S14339  
A/Molecule type: Protein  
A/Residues: 270-329 <NIE>  
R/Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A/Title: Internal homologues of the Ba fragment from human complement component factor B  
A/Reference number: A44628; MUID:84158524; PMID:6323161  
A/Accession: A44628  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 16-225/227-259 <MO>  
R/Schwable, W.; Lutting, B.; Sokolowski, T.; Esteller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A/Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A/Reference number: S14409; MUID:94041399; PMID:8225386  
A/Accession: S14409  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-764 <RES>  
A/Cross-references: GB:667310; NID:G452937; PIDN:AA13989.1; PID:G4261689  
R/Horluchi, T.; Kim, S.; Matsuno, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1567-1592, 1993  
A/Title: Human complement factor B: CDNA cloning, nucleotide sequencing, phenotypic con

A/Reference number: I57824; MUID:94067177; PMID:8247029  
A/Accession: I57824  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-31/0/33-764 <RB2>  
A/Cross-references: GB:U15702; NID:G291921; PIDN:AAA16820.1; PID:G291922  
C/Comment: 292-Cys has a free sulfhydryl.  
C/Genetics:  
A/Genes: GDB:BF  
A/Cross-references: GDB:I19726; OMIM:138470  
A/Map position: 6p21.3-6p21.3  
A/Introns: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 69  
A/Note: the list of introns may be incomplete  
A/Note: gene is located in the major histocompatibility complex, class III region  
A/Complex: complement factor B initially forms an inactive complex with complement facto  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C/Function:  
A/Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha c  
a1  
A/Pathway: complement alternate pathway  
C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; v  
C/Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydrol  
P:1-35/Domain: signal sequence #status predicted <SIG>  
P:26-764/Product: complement factor B #status experimental <MAT>  
P:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
P:37-98/Domain: complement factor H repeat homology <FH1>  
P:103-158/Domain: complement factor H repeat homology <FH2>  
P:165-218/Domain: complement factor H repeat homology <FH3>  
P:260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
P:268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
P:482-752/Domain: trypsin homology #status atypical <TRY>  
P:37-76/62-98, 103-145, 131-158, 165-205, 191-218, 478-596, 511-527, 599-615, 656-682, 695-725/DI  
F:122,142,285,378/Binding site: carbohydrate (Aan) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526,576,699/Active site: His, Asp, Ser #status experimental  
  
Query Match 11.5%; Score 111.5; DB 1; Length 764;  
Best local similarity 22.8%; Pred. No. 0.11;  
Matches 49; Conservative 44; Mismatches 77; Indels 45; Gaps 10;  
  
Qy 1 GGFPLYITLDKSGV-----LHMNELYFPEQLAHFISQLMSRIVSTRTITMLK 55  
Db 267 GSNMIVLVLDGSSIGASNFTGAKKCVNLEKVASGVXK-RYGLVYATYPIKIWKV 324  
Qy 56 TE-----DREQIRQLLEL-----OKVLPGSDPYVMEGPERSEQIYENR-----QGY- 99  
Db 325 SEADSSNADWYTKQLNINEYEDKLSGTYN-----KKALQAVYSMSMSNDDVPPGWN 378  
Qy 100 RTASVITATLDG-----ELHEDLFPYSREBNRSRDLGAIYCVG--VXDFN 144  
Db 379 RTRRVITLMTDGLHNMGGDPITVIDELRDLLYIKKRNRPBDLDVYFVGVLVQVN 438  
Qy 145 ETQLARIADSKDHYFPVNDGFQALQGIHSLTKS 179  
Db 439 INMLASKDNEQHFVKYD-MENLEDVYQWIDIS 472  
  
RESULT 13  
151027  
type XII collagen alpha-1 chain - eastern newt (fragment)  
C/Species: Nocomphthalmus viridescens, Triturus viridescens (eastern newt)  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
C/Accession: I51027  
R/Mel, Y.; Yang, E.V.; Klett, K.P.; Tassava, R.A.  
Dev. Biol. 168, 503-513, 1995  
A/Title: Monoclonal antibody WT2 identifies the urodele alpha 1 chain of type XII collagen  
A/Reference number: I51027; MUID:95246925; PMID:7729585  
A/Accession: I51027  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-929 <MEI>  
A/Cross-references: UNIPROT:Q91145; EMBL:U19494; NID:G633647; PIDN:AAA80217.1; PID:G63264  
F:155-236/Domain: fibronectin type III repeat homology <3FR>



Db 210 LVSRRVC 216

## RESULT 15

821369 collagen alpha 2(VI) chain precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: S21369; S28808; S13745; S18863

R/Author: X.Y.Z.; Bardon, S.; Dani, C.

A/Title: Cloning of alpha2 chain of type VI collagen and expression during mouse develop

A/Reference number: S28808; PMID:8380980

A/Accession: S21369

A/Molecule type: mRNA

A/Residues: 1-1029 <IBR>

A/Cross-references: UNIPROT:Q02788; EMBL:X65582; NID:G49808; PDB:CAA46541.1; PID:G49808

R/Author: A.; Bertrand, B.; Bardon, S.; Dani, E.Z.; Grimaldi, P.; Althaud, G.; Dani, C

A/Title: Cloning of alpha2 chain of type VI collagen and expression during mouse develop

A/Reference number: S28808; PMID:8380980

A/Accession: S28808

A/Molecule type: mRNA

A/Residues: 266-1029 <IB2>

A/Cross-references: EMBL:X62332; NID:G49906; PDB:CAA44206.1; PID:G49907

R/Author: R.; Constantino, C.D.; Jimenez, S.A.

Matrix 11, 1-9, 1991

A/Title: Structure of cDNAs encoding the triple-helical domain of murine alpha-2(VI) col

rate oligonucleotides for generation of novel cDNA clones.

A/Reference number: S13745; PMID:1709252

A/Accession: S13745

A/Molecule type: mRNA

A/Residues: 266-267, 'S', 269-294, 'L', 296-600 <CON>

A/Cross-references: GB:106343; NID:G192671; PDB:AAA37441.1; PID:G192672

A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 4 in having 306-116

C/Keywords: alternative splicing; cartilage; cell binding; connective tissue; disulfide

F/1-28/Domain: signal sequence #status predicted <SIG>

F/29-1029/Product: collagen alpha 2(VI) chain #status predicted <MAT>

F/29-265/Domain: globular #status predicted <NC2>

F/54-226/Domain: von Willebrand factor type A repeat homology <VWA1>

F/266-600/Domain: collagenous #status predicted <COL>

F/376-378/Region: cell attachment (R-G-D) motif

F/436-438/Region: cell attachment (R-G-D) motif

F/499-501/Region: cell attachment (R-G-D) motif

F/508-510/Region: cell attachment (R-G-D) motif

F/549-551/Region: cell attachment (R-G-D) motif

F/601-1029/Domain: globular #status predicted <NC1>

F/623-799/Domain: von Willebrand factor type A repeat homology <VWA2>

F/841-1012/Domain: von Willebrand factor type A repeat homology <VWA3>

F/150,337,640,907,963/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 11.3%; Score 109.5; DB 1; Length 1029;

Best Local Similarity 23.5%; Pred. No. 0.24; Indels 37; Gaps 8;

Matches 50; Conservative 41; Mismatches 85; Indels 37; Gaps 8;

```

QY 4 DLYFIIDKSGSV-----LHHMB-YYFVEQLAKFTISPOLRMSF-----IVFSTRG 49
DB 56 NVYFVLDTSSESVAMOSTDSILYHMOQFVQFISQLONEFLDQVALSMRYGGLHFSQV 115
QY 50 TLMKLTEDREBQIQGLLELOKVLPGGDTYMHGEPERASBOIYYENRQGYRTASVIALT 109
DB 116 EVFSPSPGSDASFTKSLQGISRFRG--TFTDCAIANMTQOIRQHVKGAVNFVAVI--T 171
QY 110 DGEIHEDLFPYSEREARSRDLGAIVYCVGV-KDFNETOLARIADSQDHVFPVN----- 162
DB 172 DGHVTSPPCGGIKQARARBEGRIRLPALAPNRNINEQGLRIANSPEHLYRNRYATMRP 231
QY 163 DGFOLQGIHSHIL-----KKSCTEI 183
DB 232 DSTBIODDTINRIKMKHAYGECYKVSCTEI 264

```

Search completed: June 13, 2005, 20:06:37  
Job time : 15.4163 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 13, 2005, 19:37:47 ; Search time 62.775 Seconds

(without alignments)  
1525.429 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_41\_227

Perfect score: 970  
Sequence: 1 GGFDFLYFLDKSGSVLHMN.....LOGIHSILKSCIEILAAE 187

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_03:.\*  
1: uniprot\_0prot:.\*  
2: uniprot\_crembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	970	100.0	564	1 ATRI_HUMAN	Q9H6X2 homo sapien
2	961	99.1	562	1 ATRI_MOUSE	Q9CZ52 mus musculu
3	550	56.7	489	1 ATRI_HUMAN	P58335 homo sapien
4	548	56.5	487	2 O6DFX2	O6DFX2 mus musculu
5	388	40.0	641	2 Q8BVW2	Q8BVW2 mus musculu
6	146	15.1	1332	2 Q9BPQ8	Q9BPQ8 halocynthia
7	141.5	14.6	1161	1 ITAD_RAT	O9QY87 rattus norv
8	141	14.5	441	2 O8T6U5	O8T6U5 mytilus edu
9	141	14.5	444	2 O8T6C3	O8T6C3 mytilus edu
10	139	14.3	3119	1 CAIC_MOUSE	O60847 mus musculu
11	138	14.2	453	2 O8T5C2	O8T5C2 mytilus gal
12	137	14.1	724	2 O04588	O04588 elmeria max
13	132.5	13.7	1152	1 ITAM_HUMAN	P11215 homo sapien
14	131	13.5	3063	1 CAIC_HUMAN	O99715 homo sapien
15	129.5	13.4	1182	1 O8C6K9	O8C6K9 mus musculu
16	129	13.3	3124	1 CAIC_CHICK	P13944 gallus galli
17	128.5	13.2	712	2 O43981	O43981 elmeria ten
18	127.5	13.1	765	2 Q9U809	Q9U809 neospora ca
19	126.5	13.0	517	2 O43853	O43853 homo sapien
20	125.5	12.9	1163	1 ITAM_HUMAN	P20702 homo sapien
21	124.5	12.8	1153	1 ITAM_MOUSE	P05555 mus musculu
22	124.5	12.8	1162	1 ITAD_HUMAN	O13349 homo sapien
23	123	12.7	1626	2 Q8NF71	Q8NF71 homo sapien
24	121.5	12.5	760	2 Q70350	Q70350 mus musculu
25	121.5	12.5	920	2 Q28984	Q28984 sus scrofa
26	118.5	12.2	760	1 CO2_MOUSE	P21180 mus musculu
27	117	12.1	440	2 O8C8T2	O8C8T2 mus musculu
28	117	12.1	848	2 O8C720	O8C720 mus musculu
29	117	12.1	1823	2 Q7PRP5	Q7PRP5 anopheles g
30	117	12.1	3567	2 Q9ES77	Q9ES77 mus musculu
31	116.5	12.0	919	2 Q75R52	Q75R52 lymanaea sta

32	116	12.0	637	2 Q8IVX1	Q8IVX1 homo sapien
33	116	12.0	1284	2 Q6P159	Q6P159 homo sapien
34	116	12.0	1329	1 KF10_HUMAN	Q9P218 homo sapien
35	116	12.0	2944	2 O63870	O63870 mus musculu
36	116	12.0	3183	2 O652C2	O652C2 caenorhabdi
37	116	12.0	3767	1 MUA3_CAEEL	P34576 caenorhabdi
38	114.5	11.8	2104	2 Q21281	Q21281 caenorhabdi
39	114.5	11.8	2104	2 Q964N4	Q964N4 caenorhabdi
40	113.5	11.7	790	2 O6DCQ6	O6DCQ6 xenopus lae
41	113	11.6	599	2 O8MVQ1	O8MVQ1 bolitena vi
42	113	11.6	1888	1 CAIE_CHICK	P32018 gallus galli
43	112.5	11.6	1337	2 O8C972	O8C972 mus musculu
44	112.5	11.6	1034	2 O8K229	O8K229 mus musculu
45	112	11.5	588	2 Q6ZED8	Q6ZED8 synchocyst

## ALIGNMENTS

RESULT 1  
ATRI\_HUMAN STANDARD: PRT; 564 AA.  
ID Q9H6X2; Q96P02; Q9NVP3;  
AC 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 05-JUN-2004 (Rel. 44, Last annotation update)  
DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
GN Name=ANTXR1; Synonyms=ATR, TEM8;  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20407466; PubMed=10947988; DOI=10.1126/science.289.5482.1197;  
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.B.,  
RA Montgomery E., Lai A., Riggs G.J., Lengauer C., Vogelstein B.,  
RA Kinzler K.W.;  
RT "Gene expressed in human tumor endothelium";  
RL Science 289:1197-1202(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH ANTHRAX TOXIN.  
RX MEDLINE=21557240; PubMed=11700562; DOI=10.1038/n510198;  
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;  
RT "Identification of the cellular receptor for anthrax toxin.";  
RL Nature 414:225-229(2001).  
RN [3]  
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A. (ISOFORM 3).  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahara K., Murakami K., Yasuda T., Watanabe M.,  
RA Shiratori K., Sudo H., Hosoi T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M.,  
RA Yamazaki M., Nishimura K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimura M., Watanabe M., Hiraoaka S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotta T.,  
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togaya S., Komai F., Hara R., Takeuchi K., Arita M.,  
RA Iiose N., Mutsaers K., Yuki H., Ohshima A., Sasaki N., Aoeuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Motiya S., Momiyama H., Satoh N., Takami S., Terashima F., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tanigami A., Fujitara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Omori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Sena T.,

RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Matanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45(2004).  
RN [4]  
RN SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rablin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.J.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska J., Smalins D.E.,  
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RN INTERACTION WITH ANTHRAX TOXIN.  
RC TISSUE=Placenta;  
RX MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
RA Scoble H.M., Ratney G.J.A., Bradley K.A., Young J.A.T.,  
RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
RT receptor";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
RN [6]  
RN SPICE ISOFORM(S) THAT ARE POTENTIAL NMD TARGET(S).  
RX PubMed=14759258; DOI=10.1186/gb-2004-5-2-r8;  
RA Hillman R.T., Green R.E., Brenner S.E.,  
RT "An unappreciated role for RNA surveillance";  
RL Genome Biol. 5:RESEARCH008.1-RESEARCH008.16(2004).  
CC -1- FUNCTION: Cellular role is not yet known.  
CC -1- SUBUNIT: Binds to the protective antigen (PA) of *Bacillus*  
CC anthracis. Binding does not occur in the presence of calcium.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Experimental confirmation may be lacking for some  
CC isoforms;  
CC Name=1;  
CC IsoId=Q9H6X2-1; Sequence=Displayed;  
CC Note=May be produced at very low levels due to a premature stop  
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
CC Name=2;  
CC IsoId=Q9H6X2-2; Sequence=VSP\_000444; VSP\_000445;  
CC Name=3;  
CC IsoId=Q9H6X2-3; Sequence=VSP\_000446; VSP\_000447;  
CC Name=4;  
CC IsoId=Q9H6X2-4; Sequence=VSP\_000448; VSP\_000449;  
CC -1- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells  
CC but not in normal endothelial cells.  
CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain.  
CC -1- SIMILARITY: Belongs to the ATR family.  
CC -1- SIMILARITY: Contains 1 VWA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF279145; AAK52094.1; -;  
DR EMBL: AF421380; AAI26496.1; -;  
DR EMBL: AK025429; BAB15128.1; ALT\_INT.  
DR EMBL: AK001463; BAA91707.1; ALT\_FRAME.  
DR EMBL: BC012074; AAI12074.1; -;  
DR Genew: HGNC:21014; ANTXR1.  
DR H-InvDB: HIX000125; -;  
DR MIM: 606410; -;  
DR InterPro: IPR008400; Anth\_Ig.  
DR InterPro: IPR008399; Ant\_C.  
DR InterPro: IPR02035; VWF\_A.  
DR Pfam: PF05587; Anth\_Ig; 1.  
DR Pfam: PF05586; Ant\_C; 1.  
DR Pfam: PF00092; VWA; 1.  
DR SMART: SM00327; VWA; 1.  
DR PROSITE: PS50234; VWF; 1.  
KM Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 32  
FT CHAIN 33 564  
FT DOMAIN 33 321  
FT TRANSMEM 322 342  
FT DOMAIN 343 564  
FT DOMAIN 44 215  
FT DOMAIN 360 368  
FT DOMAIN 506 564  
FT CARBOHYD 166 166  
FT CARBOHYD 184 184  
FT CARBOHYD 262 262  
FT VARSPPLIC 365 368  
FT FT  
FT VARSPPLIC 369 564  
FT FT  
FT VARSPPLIC 268 297  
FT FT  
FT VARSPPLIC 298 564  
FT FT  
FT VARSPPLIC 319 333  
FT FT  
FT VARSPPLIC 334 564  
FT FT  
SQ SEQUENCE 564 AA; 62789 MW; B18A00D5DF2233 CRC64;  
Query Match 100.0%; Score 970; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 2.1e-74;  
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGFDFLFIIDKSGSVLHNNHNIYFVEQLAHKRISPOLBMSFTVSRGTLTKLTRE 60  
DB 41 GGFDFLFIIDKSGSVLHNNHNIYFVEQLAHKRISPOLBMSFTVSRGTLTKLTRE 100  
QY QIRQGLSEIQKVPVPGDVTYHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFY 120  
DB 101 QIRQGLSEIQKVPVPGDVTYHGEFERASEQIYYENRQGYRTASVIALTDGELHEDLFY 160  
QY 121 SEERANRSRDGLAIYVCVGVKDFNEFQIARIADSKDHPVNDGFQALOGIHSILKSC 180  
DB 161 SEERANRSRDGLAIYVCVGVKDFNEFQIARIADSKDHPVNDGFQALOGIHSILKSC 220  
QY 181 IETLAIE 187  
DB 221 IETLAIE 227  
-----  
RESULT 2  
ATTRI\_MOUSE STANDARD; PRT; 562 AA.

CC 09c252-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
 GN Name=htxrl; Synonyms=Atx, Tem8;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Mus.  
 NC NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1)  
 RX MEDLINE=21443268; PubMed=11559528;  
 RA Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,  
 RA Kinzler K.W., St Croix B.  
 RT "Cell surface tumor endothelial markers are conserved in mice and  
 humans."  
 RL Cancer Res. 61:6649-6655 (2001).  
 RN [2]  
 RP SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=22354683; PubMed=12466651; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kanakawa T., Adachi T., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamamaki I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,  
 RA Baldevelil R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schmitt L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,  
 RA Dalle E., Dregani T.A., Fletcher C.F., Forrest A., Frizer K.S.,  
 RA Gaasterland T., Gariboldi M., Giesli C., Godzik A., Gough J.,  
 RA Grimmond S., Guelinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzielski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lemhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Milt H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Petrea G., Peocle G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sanderlin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilmig L.G., Wrynshaw-Boris A., Yamagisawa M., Yang I., Yang L.,  
 RA Yun Z., Zavolan M., Zhu Y., Zimmer A., Carrilho P., Hayatsu N.,  
 RA Hirozaki T., Ishikawa T., Kono H., Nakamura M., Sakakuma N., Sato K.,  
 RA Shirozaki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imocani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Saeki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 CC -1- FUNCTION: Cellular role is not yet known. (PA) of Bacillus  
 anthracis (By similarity).  
 CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 anthracis (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9CZ52-1; Sequence=displayed;  
 CC Name=2;  
 CC IsoId=Q9CZ52-2; Sequence=VSP 000450;  
 CC Note=No experimental confirmation available;  
 CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain  
 (By similarity).  
 CC -1- SIMILARITY: Belongs to the ATR family.  
 CC -1- SIMILARITY: Contains 1 VWA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .
CC	EMBL; AF378762; AAL11939.1; -
DR	EMBL; AK013005; BAB28591.1; ALT_INIT.
DR	MGI; MGI:1916788; Antxr1.
DR	InterPro; IPR008400; Antch_Ig.
DR	InterPro; IPR008359; Ant_C.
DR	InterPro; IPR02035; VWF_A.
DR	Pfam; PF05587; Antch_Ig; 1.
DR	Pfam; PF05586; Ant_C; 1.
DR	Pfam; PF00092; VWF; 1.
DR	PROSITE; PS00234; VWF_A; 1.
KM	Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.
FT	SIGNAL
FT	CHAIN
FT	TRANSMEM
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	VARSPLIC
FT	SEQUENCE
FT	Best Local Similarity
FT	Matches
QY	1 GGFPLYLFLDKSGSVLHHNNEIYFVEQLAKFTISPOLKNSFIYFSTRGTTMLKLTEDRE 60
DB	39 GGFDFLYFLDKSGSVLHHNNEIYFVEQLAKFTISPOLKNSFIYFSTRGTTMLKLTEDRE 98
QY	61 QIRGSELEQLKVLPGGDTYVMEGSEPRASEOITYENRQGYPRASYIALTDGELHEDLFFY 120
DB	99 QIRGSELEQLKVLPGGDTYVMEGSEPRASEOITYENRQGYPRASYIALTDGELHEDLFFY 158
QY	121 SERANRSDRDGAIVYCVGVDPFNETQARIADSKDHVPFVNDGFOALOGIHSILKSKC 180
DB	159 SERANRSDRDGAIVYCVGVDPFNETQARIADSKDHVPFVNDGFOALOGIHSILKSKC 218
QY	181 IEILAAE 187
DB	219 IEILAAE 225
RESULT 3	
ATTR2_HUMAN	STANDARD; PRT; 489 AA.
AC	P58335; Q86U11; Q8N813; Q96NC7;
DT	28-FEB-2003 (Rel. 41, Created)
DT	10-OCT-2003 (Rel. 42, Last sequence update)
DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Antitrac toxin receptor 2 precursor (Capillary morphogenesis protein-2) (CMG-2).
DE	Name=ANTXR2; Synonyms=CMG2;
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	SEQUENCE FROM N.A. (ISOFORM 2).
RP	MEDLINE=21539596; PubMed=11683410;
RX	Bell S.E., Mavila A., Salazar R., Bayless K.J., Kanagala S.,
RA	Maxwell S.A., Davis G.E.;
RT	"Differential gene expression during capillary morphogenesis in 3D collagen matrices: regulated expression of genes involved in basement

RT membrane matrix assembly, cell cycle progression, cellular  
RT differentiation and G-protein signaling.";  
RT J. Cell Sci. 114:2755-2773(2001).  
RN [2]  
RN SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH ANTHRAX TOXIN.  
RC TISSUE=Placenta;  
RX MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
RA Scobie H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.;  
RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
RT receptor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
RN [3]  
RN SEQUENCE OF 78-489 FROM N.A. (ISOFORM 3), AND SEQUENCE FROM N.A.  
RN (ISOFORM 4).  
RC TISSUE=Synovial cell;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M.,  
RA Yamazaki M., Nimomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,  
RA Iehida S., Oho Y., Takiguchi S., Watanabe S., Yoshida M., Houcha T.,  
RA Kusano Y., Kanehori K., Komai F., Hara R., Takeuchi K., Arita M.,  
RA Imose N., Mueselink K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Nomiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishiyaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Oho T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45(2004).  
CC -1- FUNCTION: Cellular role is not yet known.  
CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
CC anthracis in a divalent cation-dependent manner, with the  
CC following preference: calcium > manganese > magnesium > zinc.  
CC Seems to bind to collagen type IV and laminin.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1 and 2).  
CC Secreted (isoform 3). Isoform 1 is expressed at the cell surface  
CC while isoform 2 is predominantly expressed within the endoplasmic  
CC reticulum and not at the plasma membrane.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=1;  
CC IsoId=P58335-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=P58335-2; Sequence=VSP\_008343;  
CC Name=3;  
CC IsoId=P58335-3; Sequence=VSP\_008344, VSP\_008345;  
CC Note=No experimental confirmation available;  
CC Name=4;  
CC IsoId=P58335-4; Sequence=VSP\_008346;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Expressed in colon, heart, kidney, lung,  
CC liver, peripheral blood leukocytes, placenta, skeletal muscle,  
CC small intestine and spleen.  
CC -1- DOMAIN: Binding to PA seems to be effected through the VMA domain.  
CC -1- SIMILARITY: Belongs to the ATR family.

CC -1- SIMILARITY: Contains 1 VMA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.embnet.org/licenses/embnet.html> or send an email to [license@ebi-embnet.org](mailto:license@ebi-embnet.org)).  
CC -----  
DR EMBL: AY040326; AAK77222.1; -;  
DR EMBL: AY233452; AAB04016.1; -;  
DR EMBL: AK055636; BAB70976.1; ALT\_INT.  
DR EMBL: AK091721; BAC03731.1; -;  
DR Genbank: HGNC:21732; ANTXR2.  
DR MIM: 608041; -;  
DR InterPro: IPR008399; Ant\_C.  
DR InterPro: IPR002035; VMA\_A.  
DR Pfam: PF00586; Ant\_C; 1.  
DR Pfam: PF00092; VMA; 1.  
DR PROSITE: PS50234; VMA; 1.  
KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 33  
FT CHAIN 34 489  
FT DOMAIN 34 318  
FT TRANSMEM 319 341  
FT DOMAIN 342 489  
FT DOMAIN 44 213  
FT CARBOHYD 250 250  
FT CARBOHYD 260 260  
FT VARSPLIC 213 315  
FT VARSPLIC 290 322  
FT VARSPLIC 323 489  
FT VARSPLIC 477 489  
FT VARSPLIC 357 357  
FT CONFLICT 489 AA; 53692 MM; B9F679DB756B2B7 CRC64;  
SQ SEQUENCE  
Query Match 56.7%; Score 550; DB 1; Length 489;  
Best Local Similarity 59.9%; Pred. No. 1.3e-38;  
Matches 109; Conservative 35; Mismatches 36; Indels 2; Gaps 1;  
QY 3 FDIYFLDKSGSYLHMNIEYVEQLAKKFIPOALMSTIVSTRGTLMLKLTEDREQI 62  
DB 43 FDIYFVLDSKGSVANNMIEYVFOQLAERFVSPENMLSPVSSQATITLPLTGDRGKI 102  
QY 63 ROGLEELQKLPAGDPTMHEGFRASBOIYENRQGRKTSVIALTDEGLEDLFFYSE 122  
DB 103 SKLEEDAKRSPVGERIYHGLKLANEQI-QVAGLKTSITIALTDGKLDGVPSYAE 160  
QY 123 REANRSRLGAIYCVGVDPNETOLARIADSKDHPVNDGFQALOGIHSILKSCIE 182  
DB 161 KEAKISRSIGASVYCVGLDFEQALERLADSKQEVFPVVGQALKGIIINSIAGCTE 220  
QY 183 IL 184  
DB 221 IL 222  
RESULT 4  
ID Q6DFX2 PRELIMINARY; PRT; 487 AA.  
AC Q6DFX2; 25-OCT-2004 (TEMBLrel. 28, Created)  
DT 25-OCT-2004 (TEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TEMBLrel. 28, Last annotation update)  
DE Anthrax toxin receptor 2.

GN Name=Antxr2; (Mouse).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Krauner R.D., Collins F.S., Wagner K.H., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkin R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marisla K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Saiton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshlyuk S., Carninci P., Prange C.,  
 RA Raba S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Boeck S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunnaracne P.H.,  
 RA Richards S., Motley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Hefley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywiński M.I., Skalek U., Small D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RA Strauberg R.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC076595; AAH76555.1; -  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0004872; P: receptor activity; IEA.  
 DR InterPro: IPR008400; Antc Ig.  
 DR InterPro: IPR008399; Antc Ig.  
 DR InterPro: IPR002035; VWF A.  
 DR Pfam: PF05587; Antc Ig1 I.  
 DR Pfam: PF05586; Antc Ig1 I.  
 DR Pfam: PF00092; VWA; 1.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PSS0234; VWPA; 1.  
 KM Receptor.  
 SO SEQUENCE 487 AA; 53184 MW; 61A400D60BC8DE9 CRC64;  
 Query Match 56.5%; Score 548; DB 2; Length 487;  
 Best Local Similarity 59.9%; Pred. No. 2e-38;  
 Matches 109; Conservative 32; Mismatches 39; Indels 2; Gaps 1;  
 QY 3 PDLYFIIDKSGSVLAHNNIYFYFEQALAHKFIISQALMSPIVSTRTGTLMLKLTEDREOI 62  
 DB 43 PDLYFVLDKSGSVANNWNIYFNPHQALTRFVSPKMLSPFVSSQATLILPLTGDRYKI 102  
 QY 63 ROGIEELQKVLPGSDTYMHGSPERASQIYYBNQGRITASVILLADGELHEDLFEYSE 122  
 DB 103 GKGIEDKAVKPVGETIYHBLKLANEOT--QNGGKASITITLTDGKLDGLVPSYAE 160  
 QY 123 REANRSHDLAIVYCVGVGFNETQTLARIADSKDHPVPVNDGFOALOGIIHSLKSCIE 182  
 DB 161 NEAKSKSLASVYCVGVLFPEQQLERIDSKQGVFPVPGKFGALGIINSILAQCTE 220  
 QY 183 IL 184  
 DB 221 IL 222  
 RESULT 5  
 Q8BVM2 PRELIMINARY; PRT; 641 AA.  
 AC Q8BVM2;  
 AC Q8BVM2;

DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched  
 DE library, clone 493343011 product: hypothetical Prolin-rich region/von  
 DE Willebrand factor type A domain containing protein, full insert  
 DE sequence.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN PANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA The PANTOM Consortium;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20493974; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.132600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh A.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaueuchi S., Ikegami T., Kaishiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Matshiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai U.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Pukuda S., Furuno M., Hanagata T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai U., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK077206; BAC36683.1; -  
 DR GO: GO:0016021; C: integral to membrane; IEA.

```

DR GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR008400; Antch_19.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF05587; Antch_19; 1.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PRO0453; VMPADOMAIN.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS50234; VMA; 1.
DR Hypothetical protein.
KM SEQUENCE 641 AA; 70415 MW; 199E300730BC85E3 CRC64;
SQ
Query March 40.0%; Score 388; DB 2; Length 641;
Best Local Similarity 43.9%; Pred. No. 1.3e-24;
Matches 82; Conservative 33; Mismatches 72; Indels 0; Gaps 0;
QY 1 GGFPLYILDKSGSVLHHMNEIYFVDELAKHFIISPOLRMSFIYFSTRGTTMLKTEBRE 60
DB 73 GGFPLYILDKSGSVADNMWHIYSGFAGLVKFTNPMLRISIIYSTEAEVILPLDTSK 132
QY 61 QIRGELBELQKVLPGSDTYMHGFEFASRSEQIYENRGYRTASVITALTDELHEDLPFY 120
DB 133 EIKNSLLVLSIVSGVLTTHQKGRKANEOIRKSTLGRIVNSVYIALTDELHEDLPFY 192
QY 121 SEREANSRDLGALVYCVGVDFMETOLARIADSKDHVPVNDGFOALOGIHSILKSKC 180
DB 193 TMEAKAKRRMGALVTVGVFMYSKQOLVNIAGDPDRCFGVDEGFSALGVDPLTSKSC 252
QY 181 IEILAAE 187
DB 253 TEILSVQ 259

```

## RESULT 6

```

Q9BP08 PRELIMINARY; PRT; 1332 AA.
ID Q9BP08
AC Q9BP08;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Integrin alpha H1 precursor.
GN Name=H1TCGAL;
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Scolidobranchia; Pyuridae; Halocynthia.
OC NCBI_TaxID=7729;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Hemocyte;
RX MEDLINE=21103187; PubMed=11160215;
RA Miyazawa S., Azumi K., Nonaka M.;
RT "Cloning and characterization of integrin alpha subunits from the
RT solitary ascidian, Halocynthia roretzi.";
RL J. Immunol. 166:1710-1715 (2001).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AB048261; BAB21479.1; -.
DR HSSP; P11215; 1BHQ.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00181; Int_alpha; 5.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS50234; VMA; 1.
KM Cell adhesion; Integrin; Signal; Transmembrane.
FT SIGNAL 1 30 Potential.
FT CHAIN 31 1332 Integrin alpha H1.

```

```

SQ SEQUENCE 1332 AA; 145851 MW; 0D9108D2B05CFEAE CRC64;
Query Match 15.1%; Score 146; DB 2; Length 1332;
Best Local Similarity 24.5%; Pred. No. 0.0014;
Matches 52; Conservative 34; Mismatches 60; Indels 66; Gaps 7;
QY 2 GFDLYFLDKSGSVLHHMNEIYFVDELAKHFIISPOLRMSFIYFSTRGTTMLKTEBRE 61
DB 204 GVDVLFVLDGSSGVGNKPFVDKVMKNVT-----AKLDIGKEI 241
QY 62 IROGLEBELQKVLPG-----GDTYMHGFEFASRSEQIYENRGYRT----- 101
DB 242 VRGVGVGVYSHVYSGKSNKQYITTEHSIGFELDNFENAVDRI---QLGGTYTYGRA 298
QY 102 -----ASVITALTDELHEDLPFYSEREANSRDLGALVYCVGVDFNET 146
DB 299 LQKIRDFDDAVYIGNKQVLLTLDGQAKDKILP--NANRLNKGIATPAVGGEYDIS 356
QY 147 QLARIA---DSKDHVPVNDGFOALOGIHSI 175
DB 357 ELKLIASGTDSTURVFTVD-FGELDSIVKSL 387

```

## RESULT 7

```

ITAD RAT
ID ITAD RAT
AC Q9QYE7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor.
GN Name=Itgad;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RA O'Brien M.M., VanderVlieten M., Kilgannon P.D., Dietrich G.,
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin.";
RL Submitted (MGC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICM3 and
CC VCAM1. May play a role in the atherosclerotic process such as
CC clearing lipoproteins from plaques and in phagocytosis of blood-
CC borne pathogens, particulate matter, and senescent erythrocytes
CC from the blood (By similarity).
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D
CC associates with beta-2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- DOMAIN: The integrin I-domain (insert) is a VMA domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VMA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF021334; AAF21241.1; -.
DR HSSP; P11215; 1BHQ.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR PRINTS; PRO1185; INTEGRINA.

```

```

DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VMA; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KM Repeat; Signal; Transmembrane.
FT SIGNAL 1 19 Potential.
FT CHAIN 1 161 Integrin alpha-D.
FT DOMAIN 20 1100 Extracellular (Potential).
FT TRANSMEM 1101 1121 Potential.
FT DOMAIN 1122 1161 Cytoplasmic (Potential).
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT 88 88 FG-GAP 2.
FT REPEAT 152 334 VMA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.
FT CA_BIND 467 475 Potential.
FT CA_BIND 531 539 Potential.
FT CA_BIND 594 602 Potential.
FT SITE 1126 1130 GPCR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 89 89 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 393 393 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 734 734 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 784 784 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 907 907 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 936 936 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491984A705E CRC64;

Query Match 14.6%; Score 141.5; DB 1; Length 1161;
Best Local Similarity 27.7%; Pred. No. 0.0029;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

QY 4 DLYFIDKSGSV-LHHNNEIYFVBOIAHKFISPOLHMSPIVSTRTGTLTKTE----- 57
DB 152 DIAFLIDGSGSINORDPAQMDPFKALMGEPASTSTLFSLMQYSNILKHTFTFEFNKIL 211
QY 58 DREIRQGLELOKVLPGDPTMHSEGFERSAQIYYENRQKYRA-SVIALTDTGELHED 116
DB 212 DPQSLVDPIVOLQ-----GLTYATYGRYTWELFHSKNGSKAKKILLVITDGQKRD 266
QY 117 LFPYSR--REANRSRDIAIYVCVGVQD-FNE-TOLARI-----ADSKDHVPVNDGFOA 167
DB 267 PLEYSVDYIPADKX---GIIRYALGVGDARQEPALKEINTIGSAPQDHYFKVGN-FAA 322
QY 168 LOGIHSILK 178
DB 323 LRSIORQLEK 333

RESULT 8
Q8T6U5 PRELIMINARY; PRT; 441 AA.
AC Q8T6U5;
DT 01-JUN-2002 (TREMBLrel. 21, Last Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Proximal thread matrix protein 1 variant a.
OS Mytilus edulis (Blue mussel).

```

```

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxId=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;
RA Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric extraorganicmic
RT byssal fibers.";
RL Biomacromolecules 3:1240-1248 (2002).
DR EMBL; AF414454; AAL83537.1;
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VMA; 2.
KM Matrix protein.
SQ SEQUENCE 441 AA; 47543 MW; 881DBBD36B891D2B CRC64;

Query Match 14.5%; Score 141; DB 2; Length 441;
Best Local Similarity 25.3%; Pred. No. 0.001;
Matches 50; Conservative 46; Mismatches 76; Indels 26; Gaps 12;

QY 1 GGFDFYIDKSGSV-LHHNNEIYFVBOIAHKF-ISPO-LRMSPIVSTRTGTLTM 53
DB 238 GHADIAFVDPASSISNNNNNNYGLMKDFMKDIYDRFNKTPDGTQPAVTVFADRAIKQF 297
QY 54 KITE--DREIRQGLELOKVLPG--GDPTMHSEGFERSAQIYYENRQ---YTAIVII 106
DB 298 GLNDYSSKAIRKID--KVTPTIQTALIGDELLENARLEV-PNNNGGREEVQKVI 353
QY 107 ALTFGEI--HEDLFYSEREANRSRDIAIYVCVGV-KDENEIOLARIADSKDHVPVND 163
DB 354 LITDGNNGHKS-----PEHSSSLRKEGVIVVAIGVGTGLKSLINIASSEEVF-TTS 408
QY 164 GPQALQGIHSILKSCI 181
DB 409 SFNLSKIMENVVTLACM 426

RESULT 9
Q8T5C3 PRELIMINARY; PRT; 444 AA.
AC Q8T5C3;
DT 01-JUN-2002 (TREMBLrel. 21, Last Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Proximal thread matrix protein 1b.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxId=6550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;
RA Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric extraorganicmic
RT byssal fibers.";
RL Biomacromolecules 3:1240-1248 (2002).
DR EMBL; AY053390; AAL17973.1;
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VMA; 2.
KM Matrix protein.
SQ SEQUENCE 444 AA; 47815 MW; D2C605347450C931 CRC64;

Query Match 14.5%; Score 141; DB 2; Length 444;
Best Local Similarity 25.3%; Pred. No. 0.001;
Matches 50; Conservative 46; Mismatches 76; Indels 26; Gaps 12;

```

```
QY 1 GGFLLYITLDSKGV-----LHNNETIYFVQLAHKF--ISPO-LRMSPIVSTRTGTTLM 53
D 241 GHADIAEFVFASSSSINANNPNNGYGLMKDFMKDIDYDFRKTGPDPTQPAVWTFADRAKQF 300
QY 54 KLTE--PREQIRQGLLELOKVLPG--GDVTWHEGFERRSEQIYYENNGQ---YNTASVIT 106
D 301 GLKDYSSKAEIKKAIID--KTPPSIIGOTAIIGDLNNRLEEV-FPNRNGGRRREVQKQVI 356
QY 107 ALTDGEL--HEDLFYSREANRSRDGAIYCVGV-KDPNETQIARIADSKDHVFPVND 163
D 357 ILTDGNNNGHKS-----PEHSSLLRKSGVIVAVIGVGGLSKSLINLASEEYVF-TTS 411
QY 164 GFQALQGIHSITLKSC 181
D 412 SFNRLSKIMENVYLAACM 429

RESULT 10
CALC_MOUSE STANDARD; PRT; 3119 AA.
ID Q60847; P70322;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=Coll2a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND
RP XIIB-1).
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=96170761; PubMed=8601036;
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
RT "Primary structure of the long and short splice variants of mouse
RT collagen XII and their tissue-specific expression during embryonic
RT development."
RT Dev. Dyn. 204:432-445 (1995).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2
RP AND XIIB-2).
RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
RX MEDLINE=99348349; PubMed=10419533; DOI=10.1074/jbc.274.31.22053;
RA Kania A.M., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
RA Olsen B.R., Nishimura I.;
RT "Structural variation of type XII collagen at its carboxyl-terminal
RT NCI domain generated by tissue-specific alternative splicing."
RT J. Biol. Chem. 274:22053-22059 (1999).
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the pericellular matrix (By similarity).
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers or any combination of the various isoforms;
CC Name=XIIA-1;
CC IsoId=Q60847-1; Sequence=Displayed;
CC Name=XIIA-2; Synonyms=ER#K;
CC IsoId=Q60847-2; Sequence=VSP_001151, VSP_001152;
CC Name=XIIB-1;
CC IsoId=Q60847-3; Sequence=VSP_001150;
CC Name=XIIB-2;
CC IsoId=Q60847-4; Sequence=VSP_001150, VSP_001151, VSP_001152;
CC -1- TISSUE SPECIFICITY: Highest expression in tendons, perichondrium,
CC skin, cornea, sclera, blood vessels, and peritoneum.
CC -1- DEVELOPMENTAL STAGE: The long NC3 X1IA isoforms are predominant at
CC early stages (ED7 and 11); at later stages of development (ED15
```

```
CC and 17) the short NC3 XIIB forms become the major forms. As the
CC short NC3 forms become the major product, the long splice variant
CC continues to be expressed in several tissues, even after birth.
CC The long NCI isoforms, XIIA-1 and XIIB-1, peak in 15-day old
CC embryos and decrease in 17-day old ones. The expression of the
CC short NCI form XIIB-2 remains constant throughout late stages of
CC embryonic development (ED15 and ED17).
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end (By similarity).
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains (By
CC similarity).
CC -1- PTM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of
CC chondroitin-sulfate type (By similarity).
CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.
CC -1- SIMILARITY: Contains 18 fibronectin type III domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -1- SIMILARITY: Contains 4 WFPA domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U25652; AAA99719.1; ALT_SEQ.
DR EMBL: U57095; AAB07047.1; -.
DR HSSP: P18614; 1MHP.
DR MGD: MGI:88448; Col12a1.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR008985; Cona1 like_1ec_g1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003129; TSP_N.
DR InterPro: IPR002035; WFPA.
DR Pfam: PF01391; Collagen; 5.
DR Pfam: PF02210; TSP_N; 1.
DR Pfam: PF00092; VWA; 4.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00060; FN3; 18.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00327; VWA; 4.
DR PROSITE: PS50853; FN3; 18.
DR PROSITE: PS50234; WFPA; 4.
KW Alternative splicing; Cell adhesion; Collagen; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Repeat; Signal; Structural protein.
FT SIGNAL 1 24
FT CHAIN 25 3119
FT DOMAIN 25 112
FT DOMAIN 140 316
FT DOMAIN 333 422
FT DOMAIN 444 620
FT DOMAIN 635 723
FT DOMAIN 726 814
FT DOMAIN 817 905
FT DOMAIN 908 997
FT DOMAIN 999 1087
FT DOMAIN 1090 1179
FT DOMAIN 1203 1375
FT DOMAIN 1388 1476
FT DOMAIN 1478 1567
FT DOMAIN 1569 1656
FT DOMAIN 1660 1747
FT DOMAIN 1758 1847
FT DOMAIN 1849 1937
FT DOMAIN 1939 2028
FT DOMAIN 2030 2119
FT DOMAIN 2121 2208
FT DOMAIN 2212 2298
and 17) the short NC3 XIIB forms become the major forms. As the
short NC3 forms become the major product, the long splice variant
continues to be expressed in several tissues, even after birth.
The long NCI isoforms, XIIA-1 and XIIB-1, peak in 15-day old
embryos and decrease in 17-day old ones. The expression of the
short NCI form XIIB-2 remains constant throughout late stages of
embryonic development (ED15 and ED17).
-1- PTM: The triple-helical tail is stabilized by disulfide bonds at
each end (By similarity).
-1- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains (By
similarity).
-1- PTM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of
chondroitin-sulfate type (By similarity).
-1- SIMILARITY: Belongs to the fibril-associated collagens with
interrupted helices (FACIT) family.
-1- SIMILARITY: Contains 18 fibronectin type III domains.
-1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
-1- SIMILARITY: Contains 4 WFPA domains.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: U25652; AAA99719.1; ALT_SEQ.
DR EMBL: U57095; AAB07047.1; -.
DR HSSP: P18614; 1MHP.
DR MGD: MGI:88448; Col12a1.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR008985; Cona1 like_1ec_g1.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR003129; TSP_N.
DR InterPro: IPR002035; WFPA.
DR Pfam: PF01391; Collagen; 5.
DR Pfam: PF02210; TSP_N; 1.
DR Pfam: PF00092; VWA; 4.
DR PRINTS: PR00453; VWFADOMAIN.
DR SMART: SM00060; FN3; 18.
DR SMART: SM00210; TSPN; 1.
DR SMART: SM00327; VWA; 4.
DR PROSITE: PS50853; FN3; 18.
DR PROSITE: PS50234; WFPA; 4.
KW Alternative splicing; Cell adhesion; Collagen; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Repeat; Signal; Structural protein.
FT SIGNAL 1 24
FT CHAIN 25 3119
FT DOMAIN 25 112
FT DOMAIN 140 316
FT DOMAIN 333 422
FT DOMAIN 444 620
FT DOMAIN 635 723
FT DOMAIN 726 814
FT DOMAIN 817 905
FT DOMAIN 908 997
FT DOMAIN 999 1087
FT DOMAIN 1090 1179
FT DOMAIN 1203 1375
FT DOMAIN 1388 1476
FT DOMAIN 1478 1567
FT DOMAIN 1569 1656
FT DOMAIN 1660 1747
FT DOMAIN 1758 1847
FT DOMAIN 1849 1937
FT DOMAIN 1939 2028
FT DOMAIN 2030 2119
FT DOMAIN 2121 2208
FT DOMAIN 2212 2298
Fibronectin type-III 9.
Fibronectin type-III 10.
Fibronectin type-III 11.
Fibronectin type-III 12.
Fibronectin type-III 13.
Fibronectin type-III 14.
Fibronectin type-III 15.
Fibronectin type-III 16.
Fibronectin type-III 17.
Fibronectin type-III 18.
```



```

FT DOMAIN 2329 2501 WMA 4.
FT DOMAIN 2525 2717 TSP N-terminal.
FT DOMAIN 2456 2751 Nonhelical region (NC3).
FT DOMAIN 2752 2899 Triple-helical region (COL2) with 1
FT DOMAIN 2900 2942 imperfection.
FT DOMAIN 2943 3045 Nonhelical region (NC2).
FT DOMAIN 2943 3045 Triple-helical region (COL1) with 2
FT DOMAIN 3046 3119 imperfections.
FT SITE 866 868 Nonhelical region (NC1).
FT SITE 2784 2786 Cell attachment site (Potential).
FT SITE 2896 2898 Cell attachment site (Potential).
FT MOD_RES 2945 2945 Hydroxyproline (By similarity).
FT MOD_RES 2948 2948 Hydroxyproline (By similarity).
FT MOD_RES 2951 2951 Hydroxyproline (By similarity).
FT MOD_RES 2960 2960 Hydroxyproline (By similarity).
FT MOD_RES 2966 2966 Hydroxyproline (By similarity).
FT MOD_RES 2969 2969 Hydroxyproline (By similarity).
FT MOD_RES 2972 2972 Hydroxyproline (By similarity).
FT MOD_RES 2984 2984 Hydroxyproline (By similarity).
FT MOD_RES 3001 3001 Hydroxyproline (By similarity).
FT MOD_RES 3004 3004 Hydroxyproline (By similarity).
FT MOD_RES 3015 3015 Hydroxyproline (By similarity).
FT MOD_RES 3024 3024 Hydroxyproline (By similarity).
FT MOD_RES 3027 3027 Hydroxyproline (By similarity).
FT MOD_RES 3030 3030 Hydroxyproline (By similarity).
FT DOMAIN 869 872 Poly-Thr.
FT CARBOHYD 704 704 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 802 802 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 893 893 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 985 985 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 1769 1769 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2212 2212 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2533 2533 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2684 2684 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1190 Missing (in isoform XIIB-1 and isoform
FT XIIB-2).
FT VARSPLIC 3062 3064 /FTId=VSP_001150.
FT VARSPLIC 3065 3119 Epy -> GSG (in isoform XIIB-2 and isoform
FT VARSPLIC 3065 3119 Missing (in isoform XIIB-2 and isoform
FT XIIB-2).
SQ SEQUENCE 3119 AA; 340239 MW; 981F99C86AB3251 CRC64;

Query Match 14.3%; Score 139; DB 1; Length 3119;
Best Local Similarity 30.3%; Pred. No. 0.015;
Matches 60; Conservative 37; Mismatches 75; Indels 26; Gaps 12;

```

```

AC Q8T5C2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Proximal thread matrix protein 1.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Periomorphia; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;
RA Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric extraorganic
RT byssal fibers";
RL Biomacromolecules 3:1240-1248 (2002).
DR EMBL: AY053391; AAL17974.1; -.
DR HSSP: P20701; IMUN.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR002035; VWP_A.
DR Pfam: PF00092; VMA; 2.
DR PRINTS: PR00453; VWPADOMAIN.
DR SMART: SM00327; VMA; 2.
DR PROSITE: PS50234; WMA; 2.
KW Matrix protein.
SQ SEQUENCE 453 AA; 48784 MW; D60497F5C0C51BED CRC64;

Query Match 14.2%; Score 138; DB 2; Length 453;
Best Local Similarity 25.3%; Pred. No. 0.0019;
Matches 50; Conservative 44; Mismatches 78; Indels 26; Gaps 12;

```

```

FT DOMAIN 2329 2501 WMA 4.
FT DOMAIN 2525 2717 TSP N-terminal.
FT DOMAIN 2456 2751 Nonhelical region (NC3).
FT DOMAIN 2752 2899 Triple-helical region (COL2) with 1
FT DOMAIN 2900 2942 imperfection.
FT DOMAIN 2943 3045 Nonhelical region (NC2).
FT DOMAIN 2943 3045 Triple-helical region (COL1) with 2
FT DOMAIN 3046 3119 imperfections.
FT SITE 866 868 Nonhelical region (NC1).
FT SITE 2784 2786 Cell attachment site (Potential).
FT SITE 2896 2898 Cell attachment site (Potential).
FT MOD_RES 2945 2945 Hydroxyproline (By similarity).
FT MOD_RES 2948 2948 Hydroxyproline (By similarity).
FT MOD_RES 2951 2951 Hydroxyproline (By similarity).
FT MOD_RES 2960 2960 Hydroxyproline (By similarity).
FT MOD_RES 2966 2966 Hydroxyproline (By similarity).
FT MOD_RES 2969 2969 Hydroxyproline (By similarity).
FT MOD_RES 2972 2972 Hydroxyproline (By similarity).
FT MOD_RES 2984 2984 Hydroxyproline (By similarity).
FT MOD_RES 3001 3001 Hydroxyproline (By similarity).
FT MOD_RES 3004 3004 Hydroxyproline (By similarity).
FT MOD_RES 3015 3015 Hydroxyproline (By similarity).
FT MOD_RES 3024 3024 Hydroxyproline (By similarity).
FT MOD_RES 3027 3027 Hydroxyproline (By similarity).
FT MOD_RES 3030 3030 Hydroxyproline (By similarity).
FT DOMAIN 869 872 Poly-Thr.
FT CARBOHYD 704 704 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 802 802 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 893 893 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 985 985 O-linked (Xyl...) (Chondroitin sulfate)
FT CARBOHYD 1769 1769 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2212 2212 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2533 2533 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 2684 2684 N-linked (GlcNAc...) (Potential).
FT VARSPLIC 25 1190 Missing (in isoform XIIB-1 and isoform
FT XIIB-2).
FT VARSPLIC 3062 3064 /FTId=VSP_001150.
FT VARSPLIC 3065 3119 Epy -> GSG (in isoform XIIB-2 and isoform
FT VARSPLIC 3065 3119 Missing (in isoform XIIB-2 and isoform
FT XIIB-2).
SQ SEQUENCE 3119 AA; 340239 MW; 981F99C86AB3251 CRC64;

Query Match 14.3%; Score 139; DB 1; Length 3119;
Best Local Similarity 30.3%; Pred. No. 0.015;
Matches 60; Conservative 37; Mismatches 75; Indels 26; Gaps 12;

```

```

RESULT 12
ID 004588 PRELIMINARY; PRT; 724 AA.
AC 004588;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Major antigen homologous sequence (emp100).
OS Eimeria maxima.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93149203; PubMed=8426611; DOI=10.1016/0166-6851(93)90255-V;
RA Pasmontes L.E., Hug D., Huembelin M., Weber G.;
RT "Sequence of a major Eimeria maxima antigen homologous to the Eimeria
RT tenella microneme protein Etp100.";
RL Mol. Biochem. Parasitol. 57:171-174 (1993).
DR EMBL: M99058; AAA29076.1; -.
DR PIR: A48569; A48569.
DR HSSP: P07966; 1LST.
DR GO: GO:0008083; F: growth factor activity; IEA.
DR InterPro: IPR000762; PTN_MK.
DR InterPro: IPR000884; TSPL.

```



```
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-M
CC associates with beta-2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and
CC granulocytes.
CC -1- DOMAIN: The integrin I-domain (insert) is a VWFA domain. Integrins
CC with I-domains do not undergo intramembranous cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11b entry;
CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL, J03925, AAA59544.1, -.
DR EMBL, M18044, AAA59491.1, -.
DR EMBL, J04145, AAA59903.1, -.
DR EMBL, S52227, AAB24821.1, -.
DR EMBL, S52152, AAB24821.1, JOINED.
DR EMBL, S52153, AAB24821.1, JOINED.
DR EMBL, S52154, AAB24821.1, JOINED.
DR EMBL, S52155, AAB24821.1, JOINED.
DR EMBL, S52157, AAB24821.1, JOINED.
DR EMBL, S52159, AAB24821.1, JOINED.
DR EMBL, S52161, AAB24821.1, JOINED.
DR EMBL, S52164, AAB24821.1, JOINED.
DR EMBL, S52165, AAB24821.1, JOINED.
DR EMBL, S52169, AAB24821.1, JOINED.
DR EMBL, S52170, AAB24821.1, JOINED.
DR EMBL, S52173, AAB24821.1, JOINED.
DR EMBL, S52174, AAB24821.1, JOINED.
DR EMBL, S52180, AAB24821.1, JOINED.
DR EMBL, S52181, AAB24821.1, JOINED.
DR EMBL, S52184, AAB24821.1, JOINED.
DR EMBL, S52189, AAB24821.1, JOINED.
DR EMBL, S52191, AAB24821.1, JOINED.
DR EMBL, S52192, AAB24821.1, JOINED.
DR EMBL, S52203, AAB24821.1, JOINED.
DR EMBL, S52212, AAB24821.1, JOINED.
DR EMBL, S52213, AAB24821.1, JOINED.
DR EMBL, S52216, AAB24821.1, JOINED.
DR EMBL, S52219, AAB24821.1, JOINED.
DR EMBL, S52220, AAB24821.1, JOINED.
DR EMBL, S52221, AAB24821.1, JOINED.
DR EMBL, S52222, AAB24821.1, JOINED.
DR EMBL, S52226, AAB24821.1, JOINED.
DR EMBL, M76724, AAA58410.1, -.
DR EMBL, M84477, AAA51960.1, -.
DR PIR, A31108, RWHU1B.
DR PDB, 1A8X, Model1, @=17-1152.
DR PDB, 1BHO, X-ray, 1/2=-.
DR PDB, 1BHQ, X-ray, 1/2=-.
DR PDB, 1IDN, X-ray, 1/2=-.
DR PDB, 1IDO, X-ray, @=140-331.
DR PDB, 1JLM, X-ray, @=143-334.
DR PDB, 1MIU, X-ray, A=137-331.
DR PDB, 1MF7, X-ray, A=144-337.
DR PDB, 1N9Z, X-ray, A=140-335.
DR PDB, 1NA5, X-ray, A=144-345.
DR Genew, HGNC:6149, ITGAM.
DR MIM, 120980, -.
DR GO, GO:0080305, C:integrin complex; TAS.
DR GO, GO:007155, F:cell adhesion; TAS.
DR InterPro, IPRO00413, Integrin_alpha.
DR InterPro, IPRO02035, VWF_A.
```

```

DR Pfam: PF01839; FG-GAP; 3.
DR Pfam: PF00357; Integrin_alpha; 1.
DR Pfam: PF00092; VWA; 1.
DR PRINTS; PRO1185; INTEGRIN.A.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00191; Int. alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; WFA; 1.
DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;
KW Glycoprotein; Integrin; Magnesium; Receptor; Repeat; Signal;
KW Transmembrane.
FT SIGNAL 1 16
FT CHAIN 17 1152 Integrin alpha-M.

Query Match 13.7%; Score 132.5; DB 1; Length 1152;
Best Local Similarity 26.3%; Pred. No. 0.017;
Matches 52; Conservative 41; Mismatches 66; Indels 39; Gaps 11

QY 4 DLYPLDYSGSYL-HHNNIYFVQLAHKFSPLQMSFIYFSRGTIMKTED---- 58
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 150 DIAPFIDSGSIIIPDPFRMKFEVST-----VMEOLKSKSLIFS-----LMQYSEFRH 199
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 59 -----REQRQLLELOKTLPGDPTVMHGSPRASQIYENKQVRLA-SVIALT 109
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 200 FTPEKEFQNNPERSLVKPTQL--GRTHTATGIRKVRLELNITNGARKNAFKILVIT 257
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 110 DGELEHEDFFYSE--REANRSRDLAIYVCVGXDFNETQLAR-----IADS-KDHVFP 160
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 258 DGEKRGDPLGVEDVYPEADR---GVIRYIVGVGAPFRSEKRGELNTIASKPRDHVG 314
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 161 VNDGFQALQGIHSILKK 178
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 315 VNN-PEALKTIONQLREK 331
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 14
CALC_HUMAN STANDARD; PRT; 3063 AA.
ID AC Q99715; Q99716;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DR Collagen alpha 1(XII) chain precursor.
DS Name=COL12A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), AND PARTIAL SEQUENCE.
RX MEDLINE=97288521; Pubmed=9143499; DOI=10.1006/geno.1997.4638;
RA Gerecke D.R., Olson P.F., Koch M., Knoll J.H.M., Taylor R.,
RA Hudson D.V., Champilaud M.-F., Olsen B.R., Burgess R.B.;
RT "Complete primary structure of two splice variants of collagen XII,
RT and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)
RT collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human
RT chromosome 6q12-q13."
RL Genomics 41:236-242(1997).
CC -!- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix (By similarity).
CC -!- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers of either isoform long or isoform short or any
CC combination of isoform long and isoform short;
CC Name=Long;
CC IsoId=Q99715-1; Sequence=Displayed;
CC Name=Short;

```

CC IsoId=O99715-2; Sequence=VSP\_001149;  
 CC -1- TISSUE SPECIFICITY: Found in collagen I-containing tissues; both  
 CC short and long isoforms appear in amnion, chorion, skeletal  
 CC muscle, small intestine, and in cell culture of dermal  
 CC fibroblasts, keratinocytes and endothelial cells. Only the short  
 CC isoform is found in lung, placenta, kidney and a squamous cell  
 CC carcinoma cell line.  
 CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at  
 CC each end (By similarity).  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains (By  
 CC similarity).  
 CC -1- PTM: O-glycosylation of isoform long; glycosaminoglycan of  
 CC chondroitin sulfate type (By similarity).  
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with  
 CC interrupted helices (FACIT) family.  
 CC -1- SIMILARITY: Contains 18 fibronectin type III domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -1- SIMILARITY: Contains 4 VWFA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U73778; AAC51244.1; -;  
 CC HSSP: U73779; AAD40483.1; -;  
 CC HSSP: P18614; 1MHP.  
 CC Genew: HGNC:2188; COL12A1.  
 CC MIM: 120320; -;  
 CC GO: GO:0005595; C:collagen type XII; TAS.  
 CC GO: GO:0001501; P:skeletal development; TAS.  
 CC InterPro: IPR008160; Collagen.  
 CC InterPro: IPR008985; Cona\_like\_Iec\_g1.  
 CC InterPro: IPR003961; FN\_III.  
 CC InterPro: IPR008957; FN\_III-like.  
 CC InterPro: IPR002035; VWF\_4.  
 CC Pfam: PF01391; Collagen; 4.  
 CC Pfam: PF02210; TSP\_N; 1.  
 CC Pfam: PF02210; TSP\_N; 1.  
 CC Pfam: PF00052; VMA\_4.  
 CC PRINTS: PR00453; VWFADOMAIN.  
 CC PROSITE: PS50853; FN3; 18.  
 CC PROSITE: PS50234; VWFA; 4.  
 CC Alternative splicing; Cell adhesion; Collagen;  
 CC Direct protein sequencing; Extracellular matrix; Glycoprotein;  
 CC Hydroxylation; Repeat; Signal; Structural protein.  
 CC Hydroxylation; Repeat; Signal; Structural protein.  
 CC SIGNL: 1 24  
 FT CHAIN 25 3063 Collagen alpha 1(XII) chain.  
 FT DOMAIN 25 112 Fibronectin type-III 1.  
 FT DOMAIN 140 316 VWFA 1.  
 FT DOMAIN 333 422 Fibronectin type-III 2.  
 FT DOMAIN 440 616 VWFA 2.  
 FT DOMAIN 631 719 Fibronectin type-III 3.  
 FT DOMAIN 722 810 Fibronectin type-III 4.  
 FT DOMAIN 813 901 Fibronectin type-III 5.  
 FT DOMAIN 904 993 Fibronectin type-III 6.  
 FT DOMAIN 995 1083 Fibronectin type-III 7.  
 FT DOMAIN 1086 1175 Fibronectin type-III 8.  
 FT DOMAIN 1199 1371 VWFA 3.  
 FT DOMAIN 1384 1472 Fibronectin type-III 9.  
 FT DOMAIN 1474 1563 Fibronectin type-III 10.  
 FT DOMAIN 1565 1652 Fibronectin type-III 11.  
 FT DOMAIN 1654 1743 Fibronectin type-III 12.  
 FT DOMAIN 1752 1841 Fibronectin type-III 13.  
 FT DOMAIN 1843 1931 Fibronectin type-III 14.  
 FT DOMAIN 1933 2022 Fibronectin type-III 15.  
 FT DOMAIN 2024 2113 Fibronectin type-III 16.  
 FT DOMAIN 2115 2202 Fibronectin type-III 17.  
 FT DOMAIN 2206 2290 Fibronectin type-III 18.

FT DOMAIN 2323 2496 VWFA 4.  
 FT DOMAIN 2520 2712 TSP N-terminal.  
 FT DOMAIN 2451 2746 Nonhelical region (NC3).  
 FT DOMAIN 2747 2898 Triple-helical region (COL2) with 1  
 FT DOMAIN 2899 2941 imperfection.  
 FT DOMAIN 2942 3044 Nonhelical region (NC2).  
 FT DOMAIN 3045 3044 Triple-helical region (COL1) with 2  
 FT DOMAIN 3045 3063 imperfections.  
 FT SITE 862 864 Cell attachment site (Potential).  
 FT SITE 2779 2781 Cell attachment site (Potential).  
 FT SITE 2895 2897 Cell attachment site (Potential).  
 FT MOD\_RES 2944 2944 Hydroxyproline (By similarity).  
 FT MOD\_RES 2947 2947 Hydroxyproline (By similarity).  
 FT MOD\_RES 2950 2950 Hydroxyproline (By similarity).  
 FT MOD\_RES 2959 2959 Hydroxyproline (By similarity).  
 FT MOD\_RES 2965 2965 Hydroxyproline (By similarity).  
 FT MOD\_RES 2968 2968 Hydroxyproline (By similarity).  
 FT MOD\_RES 2971 2971 Hydroxyproline (By similarity).  
 FT MOD\_RES 2983 2983 Hydroxyproline (By similarity).  
 FT MOD\_RES 3000 3000 Hydroxyproline (By similarity).  
 FT MOD\_RES 3003 3003 Hydroxyproline (By similarity).  
 FT MOD\_RES 3014 3014 Hydroxyproline (By similarity).  
 FT MOD\_RES 3023 3023 Hydroxyproline (By similarity).  
 FT MOD\_RES 3026 3026 Hydroxyproline (By similarity).  
 FT MOD\_RES 3029 3029 Hydroxyproline (By similarity).  
 FT CARBOHYD 700 700 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 798 798 O-linked (Xyl...) (Chondroitin sulfate).  
 FT CARBOHYD 889 889 O-linked (Xyl...) (Chondroitin sulfate).  
 FT CARBOHYD 981 981 O-linked (Xyl...) (Chondroitin sulfate).  
 FT CARBOHYD 1763 1763 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 2206 2206 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 2528 2528 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 2679 2679 N-linked (GlcNAc...) (Potential).  
 FT VARSPLIC 25 1188 Missing (in isoform Short).  
 FT FTId=VSP\_001149.  
 SQ SEQUENCE 3063 AA; 333189 MW; 75FEA78FA8E48293 CRC64;  
 Query Match 13.5%; Score 131; DB 1; Length 3063;  
 Best Local Similarity 29.98%; Pred. No. 0.07; 75; Indels 26; Gaps 13;  
 Matches 59; Conservative 38; Mismatches 26; Gaps 13;  
 QY 4 DLYFIIDKSGSV-LTHMNEIYFVEQLAKHF-TSP-QLRMSFVFTST---RGTLMKLTRE 57  
 DB 440 DIVFLVDGYSIGIAFVRAFLVLSFELSPRVQSLVQYRDRPTEFTTKKFTK 499  
 QY 58 DREQIQGLAEIQKLP--GGDTYMHGFEPRASEQIYENRQGYRT--ASVITATDDEL 113  
 DB 500 VEDII-----EAIPTPYRGSGTNTGKAMTYVREKIFVPSK-GSRSNVPMILITDGK- 552  
 QY 114 HEDLFYSREANRSDDLAIYCVGVKDENEQLARIAD--SKDVFPVNDPQALOGI 171  
 DB 553 SSDAF--RDPALKLNSDVEIFAVGVKDAVNSELAIASPPAEHTFVED--FDAFQRI 608  
 QY 172 IHSILKSCIEI---LAA 186  
 DB 609 SFEIQTGICRIEQLAA 626  
 RESULT 15  
 Q8C6K9 PRELIMINARY; PRT; 1182 AA.  
 ID Q8C6K9  
 AC Q8C6K9;  
 DT 01-MAR-2003 (Tremblrel). 23, Created)  
 DT 01-MAR-2003 (Tremblrel). 23, Last sequence update)  
 DT 01-JUN-2003 (Tremblrel). 24, Last annotation update)  
 DB Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN full-  
 DE length enriched library clone:E330019B14 product:hypothetical von  
 DE Willebrand factor type A domain containing protein, full insert  
 DE sequence.



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 13, 2005, 19:37:17 ; Search time 67.0133 Seconds  
(without alignments)  
1044.623 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_42\_222  
Perfect score: 943  
Sequence: 1 GFDLYRILDKSGSVLHMNE.....DGFQALQGIIHSILKSCIE 161

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Geneseq\_16Dec04:.\*  
2: geneseqp19808:.\*  
3: geneseqp19908:.\*  
4: geneseqp20018:.\*  
5: geneseqp20028:.\*  
6: geneseqp20038:.\*  
7: geneseqp20038:.\*  
8: geneseqp20048:.\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	943	100.0	297	4 AAM38976	Aam38976 Human pol
2	943	100.0	328	7 ADI00558	ADI00558 Human TAN
3	943	100.0	328	7 ADM64584	ADM64584 Human TAN
4	943	100.0	333	3 AAB01422	Aab01422 Human TAN
5	943	100.0	333	5 ABP54905	ABP54905 Human ant
6	943	100.0	333	7 ADI00534	ADI00534 Human TAN
7	943	100.0	333	7 ADM64568	ADM64568 Human von
8	943	100.0	342	7 ADI00554	ADI00554 Human TAN
9	943	100.0	342	7 ADM64580	ADM64580 Human TAN
10	943	100.0	345	7 ADI00556	ADI00556 Human TAN
11	943	100.0	345	7 ADM64582	ADM64582 Human TAN
12	943	100.0	368	5 ABP54903	ABP54903 Human ant
13	943	100.0	384	7 ADM64586	ADM64586 TANGO197
14	943	100.0	403	4 AAB01439	Aae01439 Human gen
15	943	100.0	403	5 ABG63874	ABG63874 Human alb
16	943	100.0	403	8 ADL77139	ADL77139 Albumin f
17	943	100.0	460	7 ADI00560	ADI00560 Human TAN
18	943	100.0	479	7 ADI00564	ADI00564 Human TAN
19	943	100.0	504	7 ADI00566	ADI00566 Human TAN
20	943	100.0	529	7 ADI00568	ADI00568 Human TAN
21	943	100.0	540	7 ADI00544	ADI00544 Human TAN
22	943	100.0	540	7 ADM64588	ADM64588 TANGO197
23	943	100.0	549	7 ADI00546	ADI00546 Human TAN
24	943	100.0	549	7 ADI00542	ADI00542 Human TAN
25	943	100.0	549	7 ADI00548	ADI00548 Human TAN

26	943	100.0	549	7 ADM64590	ADM64590 TANGO197
27	943	100.0	549	7 ADM64592	ADM64592 TANGO197
28	943	100.0	551	7 ADI00550	ADI00550 Human TAN
29	943	100.0	551	7 ADM64576	ADM64576 Mouse TAN
30	943	100.0	564	5 ABB90750	ABB90750 Human Tum
31	943	100.0	564	5 ABB90724	ABB90724 Human Tum
32	943	100.0	564	5 ABP54904	ABP54904 Human ant
33	943	100.0	564	6 ABP54457	ABP54457 Human Tum
34	943	100.0	564	6 ABUS4431	ABUS4431 Human Tum
35	943	100.0	564	7 ADI00552	ADI00552 Human TAN
36	943	100.0	564	7 ADJ70017	ADJ70017 Human hea
37	943	100.0	564	7 ADM64578	ADM64578 Human TAN
38	943	100.0	564	8 ADR48216	ADR48216 Human Tum
39	938	99.5	403	5 AAB01469	Aae01469 Human gen
40	938	99.5	403	5 ABG63873	ABG63873 Human alb
41	938	99.5	403	8 ADL77138	ADL77138 Albumin f
42	934	99.0	562	5 ABB90731	ABB90731 Mouse Tum
43	934	99.0	562	5 ABB90785	ABB90785 Mouse Tum
44	934	99.0	562	5 ABUS4492	ABUS4492 Mouse Tum
45	934	99.0	562	6 ABUS4438	ABUS4438 Mouse Tum

## ALIGNMENTS

RESULT 1  
AAM38976  
ID AAM38976 standard; protein; 297 AA.  
XX  
AC AAM38976;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2121.  
XX  
KW Human, noctropic; immunosuppressant; cyostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO20015312-A1.  
XX  
PD 26-JUL-2001.  
XX  
XX  
PF 26-DEC-2000; 2000WC-US034263.  
XX  
XX  
PR 23-DEC-1999; 99US-00471275.  
XX  
PR 21-JAN-2000; 2000US-00488725.  
XX  
PR 25-APR-2000; 2000US-00552317.  
XX  
PR 20-JUN-2000; 2000US-00598042.  
XX  
PR 19-JUL-2000; 2000US-00620312.  
XX  
PR 03-AUG-2000; 2000US-00653450.  
XX  
PR 14-SEP-2000; 2000US-00662191.  
XX  
PR 19-OCT-2000; 2000US-00693036.  
XX  
PR 29-NOV-2000; 2000US-00727344.  
XX  
XX  
PA (HYSE-) HYSBO INC.  
XX  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Goodrich R, Drmanac RT;  
XX  
XX  
DR WPI; 2001-442253/47.  
XX  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
XX  
PS Example 4; SEQ ID NO 2121; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AA158642-AA162213) with neurotrophic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilization of the activities such as: Immune system suppression,  
CC Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification

XX Sequence 297 AA;

Query Match 100.0%; Score 943; DB 4; Length 297;  
Best Local Similarity 100.0%; Pred. No. 1.7e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMKLTREQ 60  
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMKLTREQ 101  
QY 61 IROGLBELQKVLPGGDTYMHGEPERASEQIYENRQGYRTASVIALTDGELHEDLPFYS 120  
DB 102 IROGLBELQKVLPGGDTYMHGEPERASEQIYENRQGYRTASVIALTDGELHEDLPFYS 161  
QY 121 ERENANRSDIGAIVYCVGVDFNFTQARIADSKDHFVPVNDGFQALOGIHSITLKSCI 180  
DB 162 ERENANRSDIGAIVYCVGVDFNFTQARIADSKDHFVPVNDGFQALOGIHSITLKSCI 221  
QY 181 E 181  
QY 222 E 222  
DB 222 E 222

#### RESULT 2

AD100558  
ID AD100558 standard; protein; 328 AA.

XX AC AD100558;

XX DT 22-APR-2004 (first entry)

XX DE Human TANGO 197 HisTag fusion protein - plasmid p0615.

XX KW fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KW cutaneous; inhalation anthrax; human; TANGO 197 HisTag fusion; mutant;  
KW plasmid p0615; mutain.

XX OS Homo sapiens.  
OS Synthetic.

XX PN US2003144193-A1.

XX PD 31-JUL-2003.

XX PF 24-JUL-2002; 2002US-00201292.

XX PR 20-DEC-2001; 2001US-00038307.

XX PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.

XX PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

XX DR MPI, 2003-720708/68.

DR N-PSDB; AD100557.

XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.

XX PS Claim 45; SEQ ID NO 26; 86pp; English.

XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 HisTag fusion protein of the invention.

XX Sequence 328 AA;

Query Match 100.0%; Score 943; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMKLTREQ 60  
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFTSPQLRMSFIVSTGTLMKLTREQ 101  
QY 61 IROGLBELQKVLPGGDTYMHGEPERASEQIYENRQGYRTASVIALTDGELHEDLPFYS 120  
DB 102 IROGLBELQKVLPGGDTYMHGEPERASEQIYENRQGYRTASVIALTDGELHEDLPFYS 161  
QY 121 ERENANRSDIGAIVYCVGVDFNFTQARIADSKDHFVPVNDGFQALOGIHSITLKSCI 180  
DB 162 ERENANRSDIGAIVYCVGVDFNFTQARIADSKDHFVPVNDGFQALOGIHSITLKSCI 221  
QY 181 E 181  
QY 222 E 222  
DB 222 E 222

#### RESULT 3

ADM64584  
ID ADM64584 standard; protein; 328 AA.

XX AC ADM64584;

XX DT 03-JUN-2004 (first entry)

XX DE Human TANGO197-His tag fusion protein #2.

XX KW antibacterial; gene therapy;  
KW von Willebrand factor A-like domain amino acid sequence;  
KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
KW inhalation anthrax; human; TANGO197; his tag; fusion protein.

XX OS Homo sapiens.  
OS Synthetic.

XX PN US2003134786-A1.

XX PD 17-JUL-2003.

XX PF 20-DEC-2001; 2001US-00038307.

XX PR 20-DEC-2001; 2001US-00038307.

XX PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.

XX PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

XX DR



DR WPI, 2003-829643/77.  
DR N-PSDB; ADM64583.  
XX  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
XX  
PS Claim 44; SEQ ID NO 26; 64bp; English.  
XX  
CC The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to Bacillus anthracis;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
CC sequence of a fusion protein comprising mature human TANGO197, thrombin  
CC cleavage site and his tag that can be used to treat exposure to or  
CC prevent a symptom of anthrax.  
XX  
XX  
SQ Sequence 328 AA;  
XX  
Query Match 100.0%; Score 943; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GFDLYFLDKSGSVLHMNNEIYFVEQLAKHFIQPMRSMFIVSTGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHMNNEIYFVEQLAKHFIQPMRSMFIVSTGTTLMKLTEDREQ 101  
XX  
QY 61 IROGLELQKVLPGSDTYMHGEPFRASEQIYENRQGRTRASVIALTDGELHEDLFFYS 120  
DB 102 IROGLELQKVLPGSDTYMHGEPFRASEQIYENRQGRTRASVIALTDGELHEDLFFYS 161  
XX  
QY 121 EREANRSRDGAIYVCVGVDPNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCT 180  
DB 162 EREANRSRDGAIYVCVGVDPNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCT 221  
XX  
QY 181 E 181  
DB 222 E 222  
XX  
XX  
RESULT 4  
AAB01422  
ID AAB01422 standard; protein; 333 AA.  
XX  
AC AAB01422;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Human TANGO 197.  
XX  
XX  
XX TANGO, 128; 140; 197; 212; 224; 239; modulating agent; asthma;  
XX graft versus-host diseases; rheumatoid arthritis; psoriasis;  
XX inflammatory bowel diseases; septic shock; ulcerative colitis;  
XX Crohn's disease; chronic myelogenous leukemia; cancer; liver disease;  
XX Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia;  
XX autoimmune disease; myasthenia gravis; autoimmune diabetes;  
XX systemic lupus erythematosus; transgenic animal; diagnosis; prognosis;  
XX prophylactic; therapeutic; human.  
XX  
XX  
XX Homo sapiens.  
XX  
XX NO200039284-A1.  
XX  
XX 06-JUL-2000.  
XX  
XX 23-DEC-1999; 99MO-US031025.  
XX  
XX 30-DEC-1998; 98US-00223546.  
XX

XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX  
XX Holtzman DA;  
XX  
XX WPI, 2000-465743/40.  
XX  
XX N-PSDB; AAA47455.  
XX  
XX  
PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224  
PT and 239 polypeptides useful for the treatment of asthma, rheumatoid  
PT arthritis, psoriasis and autoimmune diseases.  
XX  
XX  
PS Claim 8; Fig 4; 209bp; English.  
XX  
XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents  
XX for regulating cellular processes like asthma, graft versus-host  
XX diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease,  
XX septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous  
XX leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis,  
XX Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis,  
XX autoimmune diabetes and systemic lupus erythematosus. The nucleic acids  
XX are also useful for producing transgenic animals and the TANGO  
XX polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239  
XX sequences are useful in forensic biology, for diagnostic assays,  
XX prognostic assays, pharmacogenomics and for monitoring clinical trials.  
XX TANGO polypeptides are suitable for both prophylactic and therapeutic  
XX methods for treating a subject at risk of a disorder or having a disorder  
XX associated with aberrant TANGO expression. A wide range of cellular  
XX disorders can be treated  
XX  
XX  
SQ Sequence 333 AA;  
XX  
Query Match 100.0%; Score 943; DB 3; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 GFDLYFLDKSGSVLHMNNEIYFVEQLAKHFIQPMRSMFIVSTGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHMNNEIYFVEQLAKHFIQPMRSMFIVSTGTTLMKLTEDREQ 101  
XX  
QY 61 IROGLELQKVLPGSDTYMHGEPFRASEQIYENRQGRTRASVIALTDGELHEDLFFYS 120  
DB 102 IROGLELQKVLPGSDTYMHGEPFRASEQIYENRQGRTRASVIALTDGELHEDLFFYS 161  
XX  
QY 121 EREANRSRDGAIYVCVGVDPNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCT 180  
DB 162 EREANRSRDGAIYVCVGVDPNETOLARIADSKDHVPVNDGFOALQGIHSLKKSCT 221  
XX  
QY 181 E 181  
DB 222 E 222  
XX  
XX  
RESULT 5  
ABP54905  
ID ABP54905 standard; protein; 333 AA.  
XX  
AC ABP54905;  
XX  
DT 08-JAN-2003 (first entry)  
XX  
DE Human anthrax toxin receptor.  
XX  
XX  
XX Anthrax; toxin; receptor; human; antibacterial.  
XX  
XX  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
XX FT 1..27  
XX FT Peptide /label= Signal\_peptide  
XX FT Protein 28..333  
XX FT /label= Mature\_protein  
XX

PN W0200246228-A2.  
XX 13-JUN-2002.  
XX  
XX 03-OCT-2001; 2001WO-US030941.  
XX  
XX 05-DEC-2000; 2000US-0251481P.  
XX  
XX (MISC ) WISCONSIN ALUMNI RES FOUND.  
XX  
XX Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
PI WPI; 2002-713265/77.  
DR N-PSDB; ABV73883.  
XX  
XX Novel isolated polypeptide useful for identifying agent that prevents or  
PT reduces effect of anthrax toxin on host cell, for treating human or non-  
XX human animal suffering from anthrax.  
XX  
XX Claim 1; Page 42-43; 45pp; English.  
XX  
XX The present sequence is the protein sequence of a polypeptide identified  
CC as a human anthrax toxin receptor (ATR) polypeptide on the basis of  
CC identity to a newly isolated human ATR (see ABP54903). The 2 polypeptides  
CC are identical between amino acids 1-317, but differ thereafter at the C-  
CC terminus. The present polypeptide, encoded by IMAGE CLONE 4563020, has no  
CC previously known function, and there has been no prior indication that it  
CC is a complete or partial ATR. The invention provides ATR polypeptides and  
CC polynucleotides, vectors, host cells, and transgenic and knock-out  
CC animals. It also provides methods for identifying molecules that bind the  
CC ATR and which reduce the toxicity of anthrax toxin. A claimed method for  
CC treating anthrax in a human or animal involves administering an agent  
CC that inhibits binding between anthrax toxin protective antigen (PA) and  
CC ATR at a level effective to reduce the severity of anthrax. Suitable  
CC agents include the present polypeptide or a PA-binding fragment of it, a  
CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
CC nucleic acid  
XX  
SQ Sequence 333 AA;  
Query Match 100.0%; Score 943; DB 5; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GPDLYPILDKSGSVLHNMNEIYFVEQLAHKFI SPOLRMSFIYSTRGTITMKLTEDREQ 60  
DB 42 GPDLYPILDKSGSVLHNMNEIYFVEQLAHKFI SPOLRMSFIYSTRGTITMKLTEDREQ 101  
QY 61 IRGGLBELQVLPGGDTYMHGEPERASBOIYENRQGYRTASVIALTDGELHEDLFFYS 120  
DB 102 IRGGLBELQVLPGGDTYMHGEPERASBOIYENRQGYRTASVIALTDGELHEDLFFYS 161  
QY 121 EREANRSRDIGAIYVCVGDVFNETQLARIADSKDHFVPVNDGFQALOGIIHSILKKSCI 180  
DB 162 EREANRSRDIGAIYVCVGDVFNETQLARIADSKDHFVPVNDGFQALOGIIHSILKKSCI 221  
QY 181 E 181  
DB 222 E 222  
RESULT 6  
ADI00534  
ID ADI00534 standard; protein; 333 AA.  
XX  
XX ADI00534;  
AC  
XX  
XX 22-APR-2004 (first entry)  
DT  
XX  
XX Human TANGO 197 protein.  
DE  
XX  
XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;

KW cutaneous; inhalation anthrax; human; TANGO 197.  
XX  
XX Homo sapiens.  
OS  
XX US2003144193-A1.  
XX  
XX 31-JUL-2003.  
XX  
XX 24-JUL-2002; 2002US-00201292.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
PI WPI; 2003-720708/68.  
XX  
XX N-PSDB; ADI00533.  
DR  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
XX to the vWF.  
XX  
XX Claim 26; SEQ ID NO 2; 86pp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 protein of the invention.  
XX  
SQ Sequence 333 AA;  
Query Match 100.0%; Score 943; DB 7; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GPDLYPILDKSGSVLHNMNEIYFVEQLAHKFI SPOLRMSFIYSTRGTITMKLTEDREQ 60  
DB 42 GPDLYPILDKSGSVLHNMNEIYFVEQLAHKFI SPOLRMSFIYSTRGTITMKLTEDREQ 101  
QY 61 IRGGLBELQVLPGGDTYMHGEPERASBOIYENRQGYRTASVIALTDGELHEDLFFYS 120  
DB 102 IRGGLBELQVLPGGDTYMHGEPERASBOIYENRQGYRTASVIALTDGELHEDLFFYS 161  
QY 121 EREANRSRDIGAIYVCVGDVFNETQLARIADSKDHFVPVNDGFQALOGIIHSILKKSCI 180  
DB 162 EREANRSRDIGAIYVCVGDVFNETQLARIADSKDHFVPVNDGFQALOGIIHSILKKSCI 221  
QY 181 E 181  
DB 222 E 222  
RESULT 7  
ADM64568  
ID ADM64568 standard; protein; 333 AA.  
XX  
XX ADM64568;  
AC  
XX  
XX 03-JUN-2004 (first entry)  
DT  
XX  
XX Human von Willebrand factor A-like domain protein TANGO197.  
DE  
XX  
XX antibacterial; gene therapy;  
KW von Willebrand factor A-like domain amino acid sequence;  
KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;

Query Match	Best Local Similarity	Score	DB	Length
Matches 181, Conservative 0, Mismatches 0, Indels 0, Gaps 0,	100.0%	943	DB 7	333
Sequence 333 AA:				
Query	1	GFDIYFLIDKSGSVLHMHNEIYYPFEOALNKRISPOLMSPIVSTRTGTTLMKLTEDDEQ	60	
Db	42	GFDIYFLIDKSGSVLHMHNEIYYPFEOALNKRISPOLMSPIVSTRTGTTLMKLTEDDEQ	101	
Qy	61	IRQGLEELQKLVPGDDTYMHGSEFERSBOIYYENRQGYRTASVIALTDGELHEDLPFYS	120	
Db	102	IRQGLEELQKLVPGDDTYMHGSEFERSBOIYYENRQGYRTASVIALTDGELHEDLPFYS	161	
Qy	121	EREANRSHDLAIYVCGVKDFNETQLARIADSKDHVPVNDGFGALOGIHSIILKSCSI	180	
Db	162	EREANRSHDLAIYVCGVKDFNETQLARIADSKDHVPVNDGFGALOGIHSIILKSCSI	221	
Qy	181	E 181		
Db	222	E 222		
RESULT 8				
ID	AD100554	standard; protein; 342 AA.		
XX	AD100554;			
XX	AC			
XX	DT	22-APR-2004 (first entry)		
XX	Human	TANGO 197 FLAG fusion protein - plasmid p0613.		

Query	Match	Score	DB	Length	342
Query Match	100.0%	Score 943	DB 7	Length 342	
Best Local Similarity	100.0%	Pred. No. 2.1e-96			
Matches 181	Conservative	0	Mismatches	0	Indels 0
DB	1	GPDLFIIDKSGSVLHNNHNIYFPFQOLAHKPSIPQLMSPFVSESTRGTTLMKLTEDRQ	60		
DB	51	GFDFLFIIDKSGSVLHNNHNIYFPFQOLAHKPSIPQLMSPFVSESTRGTTLMKLTEDRQ	110		
DB	61	IRQGLEELQKVLPGGDTVMHGFERSBOIYVNNQGYRTASVIALTDGELHEDLFFYS	120		
DB	111	IRQGLEELQKVLPGGDTVMHGFERSBOIYVNNQGYRTASVIALTDGELHEDLFFYS	170		
DB	121	EREANRSHDLGAIYVCGVKDFNETQLARIDSKDHPVNDGFOALOGIINSHLKKSC	180		
DB	171	EREANRSHDLGAIYVCGVKDFNETQLARIDSKDHPVNDGFOALOGIINSHLKKSC	230		
DB	181	E 181			
DB	231	E 231			

KW antibacterial; gene therapy;  
KW von Willebrand factor A-like domain amino acid sequence;  
KW vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
KW Inhalation anthrax; human; TANGO197; FLAG; fusion protein; gene.  
XX  
OS Homo sapiens.  
OS Synthetic.  
PN US2003134786-A1.  
PD 17-JUL-2003.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
PF 20-DEC-2001; 2001US-00038307.  
PR 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-829643/77.  
DR N-PSDB; ADM64579.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
XX  
PS Claim 44; SEQ ID NO 22; 64bp; English.  
XX  
XX The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to Bacillus anthracis;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
CC sequence of a fusion protein comprising human TANGO197 and FLAG that can  
CC be used to treat exposure to or prevent a symptom of anthrax.  
XX  
XX  
SQ Sequence 342 AA;  
Query Match 100.0%; Score 943; DB 7; Length 342;  
Best Local Similarity 100.0%; Pred. No. 2.1e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GPDLYFTLDKSGSVLHHNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTLMKLTEDREQ 60  
DB 51 GPDLYFTLDKSGSVLHHNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTLMKLTEDREQ 110  
QY 61 IROGLELOKVLPGSDTYMHGEPERASEQIYENRQGYRTASVITALTDELHEDLFFYS 120  
DB 111 IROGLELOKVLPGSDTYMHGEPERASEQIYENRQGYRTASVITALTDELHEDLFFYS 170  
QY 121 EREANRSRDIGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQALOGIHSILKKSCI 180  
DB 171 EREANRSRDIGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQALOGIHSILKKSCI 230  
QY 181 E 181  
DB 231 E 231  
RESULT 10  
AD100556  
ID AD100556 standard; protein; 345 AA.  
XX  
AC AD100556;  
XX

DT 22-APR-2004 (first entry)  
XX  
XX Human TANGO 197 Histag fusion protein - plasmid pO614.  
DE  
XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
XX cutaneous; inhalation anthrax; human; TANGO 197 Histag fusion; mutant;  
KW plasmid pO614; mutain.  
XX  
XX Homo sapiens.  
OS Synthetic.  
OS  
FH Key Location/Qualifiers  
FT Misc-difference 331 /note="Wild-type Cys replaced by Ser followed by  
FT thrombin cleavage site and His tag"  
XX  
XX US2003144193-A1.  
PD 31-JUL-2003.  
XX  
XX 24-JUL-2002; 2002US-00201292.  
PF 24-JUL-2002; 2002US-00201292.  
PR 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI; 2003-720708/68.  
DR N-PSDB; AD100555.  
XX  
XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
XX  
PS Claim 45; SEQ ID NO 24; 86bp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 Histag fusion protein of the invention.  
XX  
XX  
SQ Sequence 345 AA;  
Query Match 100.0%; Score 943; DB 7; Length 345;  
Best Local Similarity 100.0%; Pred. No. 2.1e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GPDLYFTLDKSGSVLHHNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTLMKLTEDREQ 60  
DB 42 GPDLYFTLDKSGSVLHHNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTLMKLTEDREQ 101  
QY 61 IROGLELOKVLPGSDTYMHGEPERASEQIYENRQGYRTASVITALTDELHEDLFFYS 120  
DB 102 IROGLELOKVLPGSDTYMHGEPERASEQIYENRQGYRTASVITALTDELHEDLFFYS 161  
QY 121 EREANRSRDIGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQALOGIHSILKKSCI 180  
DB 162 EREANRSRDIGAIYVCVGDVFNETOLARIADSKDHVPVNDGFQALOGIHSILKKSCI 221  
QY 181 E 181  
DB 222 E 222  
RESULT 11

ID	ADM64582
XX	ADM64582 standard; protein; 345 AA.
AC	
XX	ADM64582;
DT	
XX	03-JUN-2004 (first entry)
DE	Human TANGO197-His tag fusion protein #1.
KW	antibacterial; gene therapy;
KV	von Willebrand factor A-like domain amino acid sequence;
RW	vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;
KW	Inhalation anthrax; human; TANGO197; his tag; fusion protein; gene.
XX	
OS	Homo sapiens.
OS	Synthetic.
PN	US2003134786-A1.
PD	17-JUL-2003.
PP	20-DEC-2001; 2001US-00038307.
PR	20-DEC-2001; 2001US-00038307.
PA	(ROTT/) ROTTMAN J B.
PA	(OKEE/) O'KEEFE T L.
PA	(OZKA/) OZKAYNAK E.
PA	(HEAL/) HEALEY J J.
PI	Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;
DR	WPJ; 2003-829643/77.
DR	N-PSDB; ADM64581.
PT	New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or
PT	ameliorating symptoms of anthrax comprises a von Willebrand factor A-like
PT	domain (vWF) amino acid sequence and an amino acid sequence heterologous
PT	to the vWF.
PS	Claim 44; SEQ ID NO 24; 64pp; English.
CC	The invention describes a fusion polypeptide comprising a von Willebrand
CC	factor A-like domain (vWF) amino acid sequence and an amino acid sequence
CC	heterologous to the vWF. Also described are: a method of preventing or
CC	ameliorating a symptom of anthrax in a subject thought to be at risk for
CC	exposure to or suspected of having been exposed to Bacillus anthracis;
CC	and a pharmaceutical composition comprising the novel fusion polypeptide.
CC	The composition and method are useful in preventing or ameliorating
CC	symptoms of cutaneous and/or inhalation anthrax. This is the amino acid
CC	sequence of a fusion protein comprising human TANGO197, thrombin cleavage
CC	site and his tag that can be used to treat exposure to or prevent a
CC	symptom of anthrax.
SO	Sequence 345 AA;
Query Match	100.0%; Score 943; DB 7; Length 345;
Best Local Similarity	100.0%; Pred. No. 2,le-96.
Matches 181; Conservative	0; Mismatches 0; Indels 0; Gaps 0
Dy	1 GFDLYFLIDKSGSVLHNNNEIYFVEQLAHKRFISPOLRMSFIVFSTRGTLTKLTEDRRQ 60
Dy	42 GFDIYFLIDKSGSVLHNNNEIYFVEQLAHKRISPOLRNSFIVFSTRGTLTKLTEDRRQ 101
Dy	61 IROGLEELQKVLPGSDTYMHSEFEASSEOITYENNGRRTASVIITLTDGEIHEDLPFFS 120
Dy	102 IROGLEELQKVLPGSDTYMHSEFEASSEOITYENNGRRTASVIITLTDGEIHEDLPFFS 161
Dy	121 EREANRSBDLGAIIVCVGVKDENEETOLARIADSKDHFVPVNDGFOLAGIIHSILKKSCI 180
Dy	162 EREANRSBDLGAIIVCVGVKDENEETOLARIADSKDHFVPVNDGFOLAGIIHSILKKSCI 221
Dy	181 E 181

Db 222 E 222

RESULT 12  
ABP54903  
ID ABP54903 standard; protein; 368 AA.  
AC ABP54903;  
DT 08-JAN-2003 (first entry)  
XX Human anthrax toxin receptor.  
XX Anthrax; toxin; receptor; human; antibacterial.  
XX Homo sapiens.FH Key Location/Qualifiers  
FT Peptide 1..27  
FT /label= Signal\_peptide  
FT Region 27..321  
FT /note= "Pe-binding fragment, begins at any amino acid in  
FT the range 27-43 and ends at any amino acid in the range  
FT 221-321, region specifically described in Claim 3"  
FT Protein 28..368  
FT /label= Mature\_protein  
FT Domain 28..320  
FT /note= "extracellular domain"  
FT 44..216  
FT /note= "von Willebrand factor A domain"  
FT Region 50  
FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
FT motif with amino acid residues 52, 54, 118 and 150"  
FT 52  
FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
FT motif with amino acid residues 50, 54, 118 and 150"  
FT 54  
FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
FT motif with amino acid residues 50, 52, 118 and 150"  
FT Region 118  
FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
FT motif with amino acid residues 50, 52, 54 and 150"  
FT 150  
FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
FT motif with amino acid residues 50, 52, 54 and 118"  
FT Domain 320..343  
FT /note= "putative transmembrane domain"  
FT 344..368  
FT /note= "cytoplasmic domain"  
W0200246228-A2.  
XX PN  
XX PD 13-JUN-2002.  
XX PF 03-OCT-2001; 2001MO-US030941.  
XX PR 05-DEC-2000; 2000US-0251481P.  
PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
XX PA  
XX DR Young JAT, Bradley KA, COLLIER RJ, MogrIDGE JS;  
PI WP1: 2002-713235/77.  
DR N-PDB; ABV73881.  
XX Novel isolated polypeptide useful for identifying agent that prevents or  
PT reduces effect of anthrax toxin on host cell, for treating human or non-  
PT human animal suffering from anthrax.  
XX  
PS Claim 1; Page 29-30; 45pp: English.  
CC The present sequence is the protein sequence of a human surface-bound

CC Anthrax toxin receptor (ATR), as predicted from an isolated cDNA clone.  
CC Anthrax toxin protective antigen (PA) binds to the ATR at a von  
CC Willebrand factor A domain located in the extracellular domain of ATR.  
CC The invention provides ATR polypeptides and polynucleotides, vectors,  
CC host cells, and transgenic and knock-out animals. It also provides  
CC methods for identifying molecules that bind the ATR and which reduce the  
CC toxicity of anthrax toxin. A claimed method for treating anthrax in a  
CC human or animal involves administering an agent that inhibits binding  
CC between PA and ATR at a level effective to reduce the severity of  
CC anthrax. Suitable agents include ATR or a PA-binding fragment of ATR, a  
CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
CC nucleic acid  
XX  
SQ Sequence 368 AA;  
Query Match 100.0%; Score 943; DB 5; Length 368;  
Best Local Similarity 100.0%; Pred. No. 2.3e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101  
QY 61 IROGLEBLQKVLPGSDTYMHGFERASEQIYYENRQGYRTASVIALTDGLHEDLFFYS 120  
DB 102 IROGLEBLQKVLPGSDTYMHGFERASEQIYYENRQGYRTASVIALTDGLHEDLFFYS 161  
QY 121 EREANRSRDIGALVYCVGVDFNETQLARIADSKDHFVPVNDGFQALQGIHSLKKSCT 180  
DB 162 EREANRSRDIGALVYCVGVDFNETQLARIADSKDHFVPVNDGFQALQGIHSLKKSCT 221  
QY 181 E 181  
DB 222 E 222  
RESULT 13  
ADM64586  
ID ADM64586 standard; protein; 384 AA.  
XX  
AC ADM64586;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE TANGO197 extracellular domain-mutant IgG Fc fusion protein #1.  
XX  
XX antibacterial; gene therapy;  
XX von Willebrand factor A-like domain amino acid sequence;  
XX VWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
XX Inhalation anthrax; human; TANGO197; immunoglobulin G; IgG;  
XX fragment of crystallisation; Fc.  
XX  
XX Homo sapiens.  
OS Synthetic.  
XX  
XX US2003134786-A1.  
XX  
XX 17-JUL-2003.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX 20-DEC-2001; 2001US-00038307.  
XX  
XX (ROTTMAN J B.  
XX (OKEE/) O'KEEFE T L.  
XX (OZKA/) OZKAYNAK E.  
XX (HEAL/) HEALEY J J.  
XX  
XX Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
XX WPI: 2003-829643/77.  
XX  
XX N-PSDB; ADM64585.  
XX  
XX

XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (VWF) amino acid sequence and an amino acid sequence heterologous  
PT to the VWF.  
XX  
XX  
PS Claim 44; SEQ ID NO 10; 64pp; English.  
XX  
XX The invention describes a fusion polypeptide comprising a von Willebrand  
XX factor A-like domain (VWF) amino acid sequence and an amino acid sequence  
XX heterologous to the VWF. Also described are: a method of preventing or  
XX ameliorating a symptom of anthrax in a subject thought to be at risk for  
XX exposure to or suspected of having been exposed to Bacillus anthracis;  
XX and a pharmaceutical composition comprising the novel fusion polypeptide.  
XX The composition and method are useful in preventing or ameliorating  
XX symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
XX sequence of a fusion protein comprising human TANGO197 extracellular  
XX region, minus the DG residues closest to the transmembrane region, and  
XX immunoglobulin G (IgG) fragment of crystallisation (Fc) with mutations  
XX L235A and G237A.  
XX  
SQ Sequence 384 AA;  
Query Match 100.0%; Score 943; DB 7; Length 384;  
Best Local Similarity 100.0%; Pred. No. 2.5e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101  
QY 61 IROGLEBLQKVLPGSDTYMHGFERASEQIYYENRQGYRTASVIALTDGLHEDLFFYS 120  
DB 102 IROGLEBLQKVLPGSDTYMHGFERASEQIYYENRQGYRTASVIALTDGLHEDLFFYS 161  
QY 121 EREANRSRDIGALVYCVGVDFNETQLARIADSKDHFVPVNDGFQALQGIHSLKKSCT 180  
DB 162 EREANRSRDIGALVYCVGVDFNETQLARIADSKDHFVPVNDGFQALQGIHSLKKSCT 221  
QY 181 E 181  
DB 222 E 222  
RESULT 14  
AAE01439  
ID AAE01439 standard; protein; 403 AA.  
XX  
AC AAE01439;  
XX  
XX 17-JUL-2001 (first entry)  
XX  
XX Human gene 4 encoded secreted protein HMLPR02, SEQ ID NO:94.  
XX  
XX Human; secreted protein; proliferative disorder; cancer; tumour;  
XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
XX inflammation; allergy; neurological disorder; Alzheimer's disease;  
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
XX cardiovascular disorder; angiogenic disorder; kidney disorder;  
XX gastrointestinal disorder; pregnancy-related disorder;  
XX endocrine disorder; infection; wound healing; vunerary; cell culture;  
XX chemotaxis; food additive; gene therapy; binding partner identification;  
XX chromosome 19.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key  
XX Peptide 1..27  
XX Protein /label=Signal\_peptide  
XX /note="Mature human secreted protein"

XX WO200134626-A1.  
XX 17-MAY-2001.  
XX 01-NOV-2000; 2000MO-US030045.  
XX 05-NOV-1999; 99US-0163581P.  
XX 30-JUN-2000; 2000US-0215133P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Komarcoullis GA, Moore PA, Birse CE, Ni J;  
XX WPI; 2001-308778/32.  
XX N-PSDB; AAD05303.  
XX New nucleic acid molecules encoding 28 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives.  
XX Claim 11; Page 485-486; 562pp; English.  
XX AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted  
CC protein genes, and AAE01436-AAE01513 represent the proteins they encode.  
CC AAE01514-AAE01544 represent human secreted protein fragments or variants.  
CC The genes and their secreted proteins are useful for preventing, treating  
CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
CC Pathological conditions can be diagnosed by determining the amount of the  
CC new protein in a sample or by determining the presence of mutations in  
CC the new genes. Specific uses are described for each of the 28 genes,  
CC based on the tissues in which they are most highly expressed, and include  
CC developing products for the diagnosis or treatment of proliferative  
CC disorders, cancer, tumors, foetal and developmental abnormalities,  
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
CC diseases (e.g., Rheumatoid arthritis), inflammation, allergies,  
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
CC pregnancy-related disorders, endocrine disorders, and infections. The  
CC proteins can also be used to aid wound healing and epithelial cell  
CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
CC before transplantation, for supporting cell culture of primary tissues,  
CC to regenerate tissues, to identify their cognate ligands or binding  
CC partners, and in chemotaxis, and can be used as a food additive or  
CC preservative to modify storage properties. Antibodies specific for a  
CC protein of the invention can be used in alleviating symptoms associated  
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
CC present sequence represents a human secreted protein of the invention  
XX  
SQ Sequence 403 AA;  
Query Match 100.0%; Score 943; DB 4; Length 403;  
Best Local Similarity 100.0%; Pred. No. 2,7e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GFDLYFLIDKSGSVLHMMNEIYFVEQLAKHFISPOLRMSFIYSTGTTLMKLTEDREQ 60  
DB 42 GFDLYFLIDKSGSVLHMMNEIYFVEQLAKHFISPOLRMSFIYSTGTTLMKLTEDREQ 101  
QY 61 IROGLEELQKVLPGGDTYVMHGEFPERASEQIYENRQGRYASVIALTDGELHEDLFFYS 120  
DB 102 IROGLEELQKVLPGGDTYVMHGEFPERASEQIYENRQGRYASVIALTDGELHEDLFFYS 161  
QY 121 EREANRSDIGAIYVCVGDQFNETQLARLADSKDHVPVNDGFQALOGIHSILKXSCI 180  
DB 162 EREANRSDIGAIYVCVGDQFNETQLARLADSKDHVPVNDGFQALOGIHSILKXSCI 221  
QY 181 E 181  
DB 222 E 222

RESULT 15  
ABG63874  
ID ABG63874 standard; protein; 403 AA.  
XX AC ABG63874;  
XX 27-AUG-2002 (first entry)  
XX DE Human albumin fusion protein #549.  
XX KM Albumin fusion protein; therapeutic protein X; human albumin; HA;  
XX KM human serum albumin; HSA; cancer; reproductive disorder;  
XX KM digestive disorder; immune disorder; endocrine disorder;  
XX KM haematopoietic disorder; neural disorder; connective disorder;  
XX KM cytoskeletal; antiinfectivity; antiinflammatory; anticancer;  
XX KM immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;  
XX KM neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
XX KM osteopathic; antiarthritic.  
XX OS Homo sapiens.  
XX OS Synthetic.  
XX PN WO200177137-A1.  
XX PD 18-OCT-2001.  
XX PF 12-APR-2001; 2001MO-US011988.  
XX PR 12-APR-2000; 2000US-0229358P.  
XX PR 25-APR-2000; 2000US-0199384P.  
XX PR 21-DEC-2000; 2000US-0256931P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Haseltine WA;  
XX WPI; 2002-010886/01.  
XX PT New fusion protein for treating disease e.g. diabetes comprises an  
PT albumin fused to a therapeutic protein.  
XX PS Claim 1; Page 874-875; 2102pp; English.  
XX The present invention relates to albumin fusion proteins comprising a  
CC therapeutic protein X and human albumin (HA, also known as human serum  
CC albumin, HSA). The proteins are useful for treating a disease or disorder  
CC that may be modulated by therapeutic protein X. The albumin extends the  
CC shelf-life of protein X, and may increase its biological in vitro/in vivo  
CC activity. The protein is useful for treating and diagnosing disorders  
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
CC disease, ulcerative colitis), immune disorders (e.g. acquired  
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,  
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,  
CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
XX  
SQ Sequence 403 AA;  
Query Match 100.0%; Score 943; DB 5; Length 403;  
Best Local Similarity 100.0%; Pred. No. 2,7e-96;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GFDLYFLIDKSGSVLHMMNEIYFVEQLAKHFISPOLRMSFIYSTGTTLMKLTEDREQ 60  
DB 42 GFDLYFLIDKSGSVLHMMNEIYFVEQLAKHFISPOLRMSFIYSTGTTLMKLTEDREQ 101  
QY 61 IROGLEELQKVLPGGDTYVMHGEFPERASEQIYENRQGRYASVIALTDGELHEDLFFYS 120  
DB 102 IROGLEELQKVLPGGDTYVMHGEFPERASEQIYENRQGRYASVIALTDGELHEDLFFYS 161

Qy	121	ERBANRSRDLGAIYYCVGVDFNETOLARIADSKOHVFPVNDGFOALOGIIHSTLKSCI	180
Db	162	ERBANRSRDLGAIYYCVGVDFNETOLARIADSKOHVFPVNDGFOALOGIIHSTLKSCI	221
Qy	181	E 181	
Db	222	E 222	

Search completed: June 13, 2005, 19:56:54  
Job time : 68.0133 secs





Db 142 DIAFLIDSGSSINQRDPAQMKDFVKALMGEPASTSTLPSLMQNSNLIKHTFTFEFNIL 201  
Qy 57 DREIIRQGLELEOKVLPFGSDTYMHEGFERASEQIYYENRQGYRTA-SVIALITDGEIHAED 115  
Db 202 DPOSIVDPVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKILLVITDGOQKRD 256  
Qy 116 LFFYSE--REANRSDLGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 166  
Db 257 PLEYSVDVTPAADKA---GIIRYALGVDAFOEPALKEIWTIGSAPPQDHVFKVGN-FAA 312  
Qy 167 LOGIHSILKK 177  
Db 313 LRSIORQLOEK 323

RESULT 2  
US-08-485-618-37  
; Sequence 37, Application US/08485618  
; Patent No. 5728533  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vlieten, Monica  
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 103  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/485,618  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,652  
; FILING DATE: 21-DEC-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32797  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-485-618-37  
Query Match 15.0%; Score 141.5; DB 1; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.2e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 3 DLVFLIDKSGSV-LHMHNEIYVFEOLAHKFIQPOLMSFIVSTGTMLKTE----- 56  
Db 142 DIAFLIDSGSSINQRDPAQMKDFVKALMGEPASTSTLPSLMQNSNLIKHTFTFEFNIL 201

Qy 57 DREIIRQGLELEOKVLPFGSDTYMHEGFERASEQIYYENRQGYRTA-SVIALITDGEIHAED 115  
Db 202 DPOSIVDPVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKILLVITDGOQKRD 256  
Qy 116 LFFYSE--REANRSDLGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 166  
Db 257 PLEYSVDVTPAADKA---GIIRYALGVDAFOEPALKEIWTIGSAPPQDHVFKVGN-FAA 312  
Qy 167 LOGIHSILKK 177  
Db 313 LRSIORQLOEK 323

RESULT 3  
US-08-362-652-37  
; Sequence 37, Application US/08362652  
; Patent No. 5766850  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, W. Michael  
; APPLICANT: Van der Vlieten, Monica  
; TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive, 6300 Sear Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,652  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/173,497  
; FILING DATE: 23-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/286,889  
; FILING DATE: 5-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams Jr., Joseph A.  
; REGISTRATION NUMBER: 38,659  
; REFERENCE/DOCKET NUMBER: 27866/32391  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-362-652-37

Query Match 15.0%; Score 141.5; DB 1; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.2e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
Qy 3 DLVFLIDKSGSV-LHMHNEIYVFEOLAHKFIQPOLMSFIVSTGTMLKTE----- 56  
Db 142 DIAFLIDSGSSINQRDPAQMKDFVKALMGEPASTSTLPSLMQNSNLIKHTFTFEFNIL 201  
Qy 57 DREIIRQGLELEOKVLPFGSDTYMHEGFERASEQIYYENRQGYRTA-SVIALITDGEIHAED 115  
Db 202 DPOSIVDPVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKILLVITDGOQKRD 256  
Qy 116 LFFYSE--REANRSDLGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 166

Db 257 PLEYSVDIIPADKA---GIIRYALGVDAFOEPALKEMLNTIGSAPPQDHVFKGN-FAA 312  
QY 167 LOGIHSILKK 177  
Db 313 LRSTORQLQEK 323

## RESULT 4

US-08-605-672-37  
Sequence 37, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-37

Query Match 15.0%; Score 141.5; DB 2; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.2e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
QY 3 DLYEILDKSGSV-LHMHNEIYFVEOLAHKRISPOLMSFIVFSTRTGTLMLKTE----- 56  
Db 142 DIAPILGSSGSIINRDPAQKMDFKALMGEPASTSTLFSLMQYSNLIKTHFTFEFKNIL 201  
QY 57 DREQIRGEBELQVLPFGSDTYHMEGPERASEQIYENRQGYRTA-SVIALTLTGELHED 115  
Db 202 DPGSLVDPIVQLQ-----GLYTTATGIRTWBELFHSKNGSRKSAKKILLVITDGQKRD 256  
QY 116 LFFYSE--REANRBRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHFPVNDGFOA 166  
Db 257 PLEYSVDIIPADKA---GIIRYALGVDAFOEPALKEMLNTIGSAPPQDHVFKGN-FAA 312

QY 167 LOGIHSILKK 177  
Db 313 LRSTORQLQEK 323

## RESULT 5

US-08-482-293A-37  
Sequence 37, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-37

Query Match 15.0%; Score 141.5; DB 2; Length 1151;  
Best Local Similarity 27.7%; Pred. No. 4.2e-07;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;  
QY 3 DLYEILDKSGSV-LHMHNEIYFVEOLAHKRISPOLMSFIVFSTRTGTLMLKTE----- 56  
Db 142 DIAPILGSSGSIINRDPAQKMDFKALMGEPASTSTLFSLMQYSNLIKTHFTFEFKNIL 201  
QY 57 DREQIRGEBELQVLPFGSDTYHMEGPERASEQIYENRQGYRTA-SVIALTLTGELHED 115  
Db 202 DPGSLVDPIVQLQ-----GLYTTATGIRTWBELFHSKNGSRKSAKKILLVITDGQKRD 256  
QY 116 LFFYSE--REANRBRDGAIVYCVGVD-FNE-TQLARI-----ADSKDHFPVNDGFOA 166  
Db 257 PLEYSVDIIPADKA---GIIRYALGVDAFOEPALKEMLNTIGSAPPQDHVFKGN-FAA 312  
QY 167 LOGIHSILKK 177

Db 313 LRSTGRLQEK 323

```
RESULT 6
US-08-943-363-37
; Sequence 37, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-37
```

```
Query Match 15.0%; Score 141.5; DB 2; Length 1151;
Best Local Similarity 27.7%; Pred. No. 4.2e-07;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

QY 3 DLYFIDKSGSV-LHHNNEIYFVEQLAHKFIPOILMSFIVSTRTTLMKLT----- 56
Db 142 DIAFLIDGSGSINQRFQAKMDFFKALMGEPASTLFSIMQYSNLIKHTFTFEFNIL 201
QY 57 DREQIRGLBELQKVLPGGDTYHGEPRASSEQIYYENRQGYRTA-SVIALTDGELHED 115
Db 202 DPOSIVDPVQLQ-----GLTYTATGIRTVWELFHSKNGSRKSAKKILLVITDGOQYRD 256
QY 116 LFPYSE--REANRSRDIAIYVCVGD-FNE-TOLARI-----ADSKDHVPVNDGFOA 166
Db 257 PLEYSDVTPAADKA---GIIRYALGVGDARFQEPFALKELNTIGSAPPQDHVFKVGN-FAA 312
QY 167 LOGIHSILKK 177
Db 313 LRSTGRLQEK 323
```

```
RESULT 7
US-09-193-043-37
; Sequence 37, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193,043
; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
; US-09-193-043-37
```

```
Query Match 15.0%; Score 141.5; DB 3; Length 1151;
Best Local Similarity 27.7%; Pred. No. 4.2e-07;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

QY 3 DLYFIDKSGSV-LHHNNEIYFVEQLAHKFIPOILMSFIVSTRTTLMKLT----- 56
Db 142 DIAFLIDGSGSINQRFQAKMDFFKALMGEPASTLFSIMQYSNLIKHTFTFEFNIL 201
QY 57 DREQIRGLBELQKVLPGGDTYHGEPRASSEQIYYENRQGYRTA-SVIALTDGELHED 115
Db 202 DPOSIVDPVQLQ-----GLTYTATGIRTVWELFHSKNGSRKSAKKILLVITDGOQYRD 256
QY 116 LFPYSE--REANRSRDIAIYVCVGD-FNE-TOLARI-----ADSKDHVPVNDGFOA 166
Db 257 PLEYSDVTPAADKA---GIIRYALGVGDARFQEPFALKELNTIGSAPPQDHVFKVGN-FAA 312
QY 167 LOGIHSILKK 177
Db 313 LRSTGRLQEK 323
```

```
RESULT 8
US-09-688-307A-37
; Sequence 37, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vlieten, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
```





Db 152 DIAPIDSGSSINORDPFAQMDPVKALMBEFASTSTVLSLMQYSNLKHFTEERKNLL 211  
 Qy 57 DREQIRQGLEIBLOKVLPGSDTYMHGEPERASEQIYENRQGYRTA-SVIALTDGELHED 115  
 Db 212 DPQSLVDPIVQLQ-----GLTTTANGIRKVMELPFSKNGSRKSAGKILLVITDQGRKD 266  
 Qy 116 LFPFSE--REANRSHDLGAIIVYCVGKD-FNE-TQLARI-----ADSKDHVPVNDGPOA 166  
 Db 267 PLEISDVLPADMDA---GIRTAIVGVDAPQEBTALKELNTIGSAPPODHVFKNV-FAA 322  
 Qy 167 LOGIHSILKK 177  
 Db 323 LRSTOROLOEK 333

```

1      RESULT 13
2      US-08-482-293A-55
3      Sequence 55, Application US/08482293A
4      Patent No. 5831029
5      GENERAL INFORMATION:
6      APPLICANT: Gallatin, W. Michael
7      APPLICANT: Van der Vieren, Monica
8      TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
9      NUMBER OF SEQUENCES: 103
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Marzshall, O'Toole, Gerstein, Murray & Borun
12     STREET: 233 South Wacker Drive, 6300 Sear Tower
13     CITY: Chicago
14     STATE: Illinois
15     COUNTRY: United States
16     ZIP: 60606-6402
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: Patent Release #1.0, Version #1.25
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/482,293A
24     FILING DATE:
25     CLASSIFICATION: 530
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER: US 08/173,497
28     FILING DATE: 23-DEC-1993
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: US 08/286,889
31     FILING DATE: 5-AUG-1994
32     PRIOR APPLICATION DATA:
33     APPLICATION NUMBER: US 08/362,652
34     FILING DATE: 21-DEC-1994
35     ATTORNEY/AGENT INFORMATION:
36     NAME: Williams Jr., Joseph A.
37     REGISTRATION NUMBER: 38,659
38     REFERENCE/DOCKET NUMBER: 27866/32684
39     TELECOMMUNICATION INFORMATION:
40     TELEPHONE: 312-474-6300
41     TELEFAX: 312-474-0448
42     TELEX: 25-3856
43     INFORMATION FOR SEQ ID NO: 55:
44     SEQUENCE CHARACTERISTICS:
45     LENGTH: 1161 amino acids
46     TYPE: amino acid
47     TOPOLOGY: linear
48     MOLECULE TYPE: protein
49     US-08-482-293A-55
50
51     Query Match      15.0%; Score 141.5; DB 2; Length 1161;
52     Best Local Similarity 27.7%; Pred. No. 4.2e-07;
53     Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;
54
55     3 DLYVILDKSGGV-LHHMNEIYYFVQLAHKPFISPOLAKSFIVFSRGTTLNKLTE----- 56
56     |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
57     152 DIAPILDDSGGGINORDPQMDPVALMGEPASTSTLPSLMQYSILKTHFTTFEKNIL 211

```

```

QY 57 DREGROGLEBELOKTLPGGDPTMHEGFERASQIYYENRGVRLA-SVIALTLDEGLHD 115
Db 212 DPOSIVDPIVDLQ-----GLTYTATGIRTWELTHSKXSKRSKAKILLVITDQKXRD 266
QY 116 LFPYSE--REANRSNDLGAIVYCVGKD-FNE-TOLARI-----ADSKDHVFPVNGPOA 166
Db 267 PLEISDVIYPADKA---GIIRYAIVGAGDAFOEPTLKLINTIGSAPPDHVERKVN-FAA 322
QY 167 LOGIITHSILKT 177
Db 323 LRSTOROLOEK 333

```

```

1      RESULT 14
2      US-08-943-363-55
3      : Sequence 55, Application US/08943363
4      : Patent No.5837478
5      : GENERAL INFORMATION:
6      : APPLICANT: Gallatin, W. Michael
7      : APPLICANT: Van der Vieren, Monica
8      : TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
9      : NUMBER OF SEQUENCES: 114
10     : CORRESPONDENCE ADDRESS:
11     : ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borum
12     : STREET: 233 South Wacker Drive, 6300 Seear Tower
13     : CITY: Chicago
14     : STATE: Illinois
15     : COUNTRY: United States
16     : ZIP: 60606-6402
17     : COMPUTER READABLE FORM:
18     : MEDIUM TYPE: Floppy disk
19     : COMPUTER: IBM PC compatible
20     : OPERATING SYSTEM: PC-DOS/MS-DOS
21     : SOFTWARE: Patentin Release #1.0, Version #1.25
22     : CURRENT APPLICATION DATA:
23     : APPLICATION NUMBER: US/08/943,363
24     : FILING DATE:
25     : CLASSIFICATION: 530
26     : PRIOR APPLICATION DATA:
27     : APPLICATION NUMBER: US 08/173,497
28     : FILING DATE: 23-DEC-1993
29     : PRIOR APPLICATION DATA:
30     : APPLICATION NUMBER: US 08/286,889
31     : FILING DATE: 5-AUG-1994
32     : PRIOR APPLICATION DATA:
33     : APPLICATION NUMBER: US 08/362,652
34     : FILING DATE: 21-DEC-1994
35     : ATTORNEY/AGENT INFORMATION:
36     : NAME: Williams Jr., Joseph A.
37     : REGISTRATION NUMBER: 38,659
38     : REFERENCE/DOCKET NUMBER: 27866/32684
39     : TELECOMMUNICATION INFORMATION:
40     : TELEPHONE: 312-474-6300
41     : TELEFAX: 312-474-0448
42     : TELEX: 25-3856
43     : INFORMATION FOR SEQ ID NO: 55:
44     : SEQUENCE CHARACTERISTICS:
45     : LENGTH: 1161 amino acids
46     : TYPE: amino acid
47     : TOPOLOGY: linear
48     : MOLECULE TYPE: protein
49     : US-08-943-363-55
50
51     Query Match 15.0%; Score 141.5; DB 2; Length 1161;
52     Best Local Similarity 27.7%; Fred. No. 4.2e-07;
53     Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10
54
55     3 DLYFLIDSGSV-LHHNMEIYFVEQLAKRISPOLRMSFIIVSTRGTLNKLTE-----56
56     ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
57     152 DIATLIDSSGSIINGRDPAQMDPFKALMGEPASTSTLFSLMQYENILKTHFTPEFKIIL 211
58     57 DREGIIRGLEELQKVLPGGDITYMEHGFEPRASEQIYENRQGRIRL-SVIALTLDELGHED 115

```

```

Db      212 DPGSLVDPPIVQLQ-----GLTYTATGIRTYMEELFHSKNGSRKSAKKILVITTDGQKYRD 266
QY      116 LFFYSE--REANRSRDIGAIVYCVGKVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 166
Db      267 PLEYSVDVTPADKA---GIIRYAIGVGDAFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 322
QY      167 LOGIHSILKK 177
Db      323 LRSIQRLQLEK 333

```

RESULT 15

```

US-09-193-043-55
; Sequence 55, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vliet, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193, 043
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173, 497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286, 889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362, 652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943, 363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

```

```

Query Match      15.0%; Score 141.5; DB 3; Length 1161;
Best Local Similarity 27.7%; Pred. No. 4.2e-07;
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

```

```

QY      3 DLYFIIDKSGSV-LHMHNEIYFVEQLAHKEFISPOLMSFIVSTRGTTMLKITE----- 56
Db      152 DIAFLIDSGSINQRDPAQMKDFVKALMGEPASTSTLFSLMQYSNLIKHTFTTFEKNIL 211
QY      57 DREQIRQGLEELQKVLPGDPTVMHEGPERASEOIYENRQGYRTA-SVITALTGDELHED 115
Db      212 DPGSLVDPPIVQLQ-----GLTYTATGIRTYMEELFHSKNGSRKSAKKILVITTDGQKYRD 266
QY      116 LFFYSE--REANRSRDIGAIVYCVGKVD-FNE-TQLARI-----ADSKDHVPVNDGFOA 166
Db      267 PLEYSVDVTPADKA---GIIRYAIGVGDAFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 322
QY      167 LOGIHSILKK 177
Db      323 LRSIQRLQLEK 333

```

Search completed: June 13, 2005, 20:05:10  
 Job time : 18.6748 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 20:03:29 ; Search time 58.0354 Seconds  
(without alignments)  
1195.537 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_42\_222  
Perfect score: 943  
Sequence: 1 GFDLYFTLDKSGSVLHMHNE.....DGFQALOGIHSILKSCIE 181

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 38334425 residues  
Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /cgnt2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*  
2: /cgnt2\_6/ptodata/1/pubppaa/PCTU\_NEW\_PUB.pep.\*  
3: /cgnt2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*  
4: /cgnt2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*  
5: /cgnt2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*  
6: /cgnt2\_6/ptodata/1/pubppaa/PCTU\_PUBCOMB.pep.\*  
7: /cgnt2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*  
8: /cgnt2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*  
9: /cgnt2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*  
10: /cgnt2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*  
11: /cgnt2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*  
12: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
13: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
14: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
15: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
16: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
17: /cgnt2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*  
18: /cgnt2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*  
19: /cgnt2\_6/ptodata/1/pubppaa/US11\_PUBCOMB.pep.\*  
20: /cgnt2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep.\*  
21: /cgnt2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*  
22: /cgnt2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	943	100.0	328	14	US-10-038-307-26
2	943	100.0	328	14	US-10-201-292-26
3	943	100.0	333	10	US-09-796-753-12
4	943	100.0	333	14	US-10-038-307-2
5	943	100.0	333	14	US-10-201-292-2
6	943	100.0	342	14	US-10-038-307-22
7	943	100.0	342	14	US-10-201-292-22
8	943	100.0	345	14	US-10-038-307-24
9	943	100.0	345	14	US-10-201-292-24
10	943	100.0	403	11	US-09-833-245-621
11	943	100.0	460	14	US-10-201-292-28

12	943	100.0	479	14	US-10-201-292-32	Sequence 32, App1
13	943	100.0	504	14	US-10-201-292-34	Sequence 34, App1
14	943	100.0	529	14	US-10-201-292-36	Sequence 36, App1
15	943	100.0	551	14	US-10-038-307-18	Sequence 18, App1
16	943	100.0	551	14	US-10-201-292-18	Sequence 18, App1
17	943	100.0	564	10	US-09-818-715-187	Sequence 187, App
18	943	100.0	564	10	US-09-818-715-232	Sequence 232, App
19	943	100.0	564	14	US-10-038-307-20	Sequence 20, App1
20	943	100.0	564	14	US-10-201-292-20	Sequence 20, App1
21	943	100.0	564	14	US-10-301-822-199	Sequence 199, App
22	943	100.0	564	16	US-10-408-765A-1823	Sequence 1823, App
23	943	100.0	564	16	US-10-474-794-187	Sequence 187, App
24	943	100.0	564	16	US-10-474-794-232	Sequence 232, App
25	938	99.5	403	11	US-09-833-245-620	Sequence 620, App
26	934	99.0	562	10	US-09-818-715-194	Sequence 194, App
27	934	99.0	562	10	US-09-818-715-301	Sequence 301, App
28	934	99.0	562	16	US-10-474-794-194	Sequence 194, App
29	934	99.0	562	16	US-10-474-794-301	Sequence 301, App
30	933	98.9	460	14	US-10-201-292-30	Sequence 30, App1
31	922	97.8	534	14	US-10-038-307-12	Sequence 12, App1
32	922	97.8	534	14	US-10-201-292-12	Sequence 12, App1
33	922	97.8	543	14	US-10-038-307-10	Sequence 10, App1
34	922	97.8	543	14	US-10-038-307-14	Sequence 14, App1
35	922	97.8	543	14	US-10-038-307-16	Sequence 16, App1
36	922	97.8	543	14	US-10-201-292-10	Sequence 10, App1
37	922	97.8	543	14	US-10-201-292-14	Sequence 14, App1
38	922	97.8	543	14	US-10-201-292-16	Sequence 16, App1
39	900	95.4	538	13	US-10-047-542-99	Sequence 99, App1
40	542	57.5	488	10	US-09-796-753-52	Sequence 52, App1
41	542	57.5	488	14	US-10-038-307-6	Sequence 6, App1
42	542	57.5	488	14	US-10-201-292-6	Sequence 6, App1
43	542	57.5	488	14	US-10-368-087-16	Sequence 16, App1
44	542	57.5	488	15	US-10-104-047-2639	Sequence 2639, App
45	540	57.3	487	10	US-09-796-753-54	Sequence 54, App1

## ALIGNMENTS

RESULT 1  
US-10-038-307-26  
Sequence 26, Application US/10038307  
Publication No. US20030134786A1  
GENERAL INFORMATION:  
APPLICANT: Theresa B. ROTTMAN  
APPLICANT: James B. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038.307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 328  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-26

Query Match	100.0%	Score 943	DB 14	Length 328
Best Local Similarity	100.0%	Pred. No. 1.6e-90		
Matches 181	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	GFDLYFTLDKSGSVLHMHNEIYVEQLAKHFIISPOLRMSTIVSTGTTIMKTEDREQ	60	
DB	42	GFDLYFTLDKSGSVLHMHNEIYVEQLAKHFIISPOLRMSTIVSTGTTIMKTEDREQ	101	
QY	61	IRQGLEELQKVLPGSGDVTYMHGFERASEQIYENRQGRASVITATDGLHDDLPFFS	120	
DB	102	IRQGLEELQKVLPGSGDVTYMHGFERASEQIYENRQGRASVITATDGLHDDLPFFS	161	
QY	121	ERANRSRDIGAIYVCVVDNFNETQARIADSKDHVPVNDGFOALOGIHSILKSCIE	180	

```
Db      162 EREANRSDLGAIYCVGVDPFNETOLARIADSKDHVPVNDGFQALOGIIHSLKKSCT 221
      181 E 181
      222 E 222

RESULT 2
US-10-201-292-26
; Sequence 26, Application US/10201292
; Publication No. US20030144193A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Englin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/201,292
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-292-26

Query Match      100.0%; Score 943; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 1,6e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFDLYFLDKSGSVLHMHMEIYYFVEQLAHKFTSPQLRMSFIVFSTRTGLMKLTEDREQ 60
      42 GFDLYFLDKSGSVLHMHMEIYYFVEQLAHKFTSPQLRMSFIVFSTRTGLMKLTEDREQ 101
Db      102 IROGLEELQVLPFGDITVMEHGFERASEQIYYENRQGYRTASVITALTDELHEDLFFYS 120
      102 IROGLEELQVLPFGDITVMEHGFERASEQIYYENRQGYRTASVITALTDELHEDLFFYS 161
Qy      121 EREANRSDLGAIYCVGVDPFNETOLARIADSKDHVPVNDGFQALOGIIHSLKKSCT 180
      162 EREANRSDLGAIYCVGVDPFNETOLARIADSKDHVPVNDGFQALOGIIHSLKKSCT 221
Db      181 E 181
      222 E 222

RESULT 3
US-09-796-753-12
; Sequence 12, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
```

```
      PRIOR FILING DATE: 1999-05-14
      PRIOR APPLICATION NUMBER: 09/336,536
      PRIOR FILING DATE: 1999-06-18
      PRIOR APPLICATION NUMBER: 09/342,687
      PRIOR FILING DATE: 1999-06-29
      PRIOR APPLICATION NUMBER: 09/345,464
      PRIOR FILING DATE: 1999-06-30
      PRIOR APPLICATION NUMBER: 09/365,164
      PRIOR FILING DATE: 1999-07-30
      PRIOR APPLICATION NUMBER: 09/399,723
      PRIOR FILING DATE: 1999-09-20
      PRIOR APPLICATION NUMBER: 09/409,634
      PRIOR FILING DATE: 1999-09-30
      PRIOR APPLICATION NUMBER: 09/471,179
      PRIOR FILING DATE: 1999-12-23
      PRIOR APPLICATION NUMBER: 09/474,071
      PRIOR FILING DATE: 1999-12-29
      PRIOR APPLICATION NUMBER: 09/474,072
      PRIOR FILING DATE: 1999-12-29
      PRIOR APPLICATION NUMBER: 09/514,010
      PRIOR FILING DATE: 2000-02-25
      PRIOR APPLICATION NUMBER: 09/516,745
      PRIOR FILING DATE: 2000-03-01
      PRIOR APPLICATION NUMBER: 09/572,002
      PRIOR FILING DATE: 2000-05-14
      PRIOR APPLICATION NUMBER: 09/597,993
      PRIOR FILING DATE: 2000-06-19
      PRIOR APPLICATION NUMBER: 09/599,596
      PRIOR FILING DATE: 2000-06-22
      PRIOR APPLICATION NUMBER: 09/630,334
      PRIOR FILING DATE: 2000-07-31
      PRIOR APPLICATION NUMBER: 09/606,565
      PRIOR FILING DATE: 2000-06-29
      PRIOR APPLICATION NUMBER: 09/606,317
      PRIOR FILING DATE: 2000-06-29
      PRIOR APPLICATION NUMBER: 09/665,666
      PRIOR FILING DATE: 2000-09-20
      PRIOR APPLICATION NUMBER: 09/677,751
      PRIOR FILING DATE: 2000-09-30
      NUMBER OF SEQ ID NOS: 162
      SEQ ID NO 12
      LENGTH: 333
      TYPE: PRT
      ORGANISM: Homo sapiens
US-09-796-753-12

Query Match      100.0%; Score 943; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 1,6e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GFDLYFLDKSGSVLHMHMEIYYFVEQLAHKFTSPQLRMSFIVFSTRTGLMKLTEDREQ 60
      42 GFDLYFLDKSGSVLHMHMEIYYFVEQLAHKFTSPQLRMSFIVFSTRTGLMKLTEDREQ 101
Db      102 IROGLEELQVLPFGDITVMEHGFERASEQIYYENRQGYRTASVITALTDELHEDLFFYS 120
      102 IROGLEELQVLPFGDITVMEHGFERASEQIYYENRQGYRTASVITALTDELHEDLFFYS 161
Qy      121 EREANRSDLGAIYCVGVDPFNETOLARIADSKDHVPVNDGFQALOGIIHSLKKSCT 180
      162 EREANRSDLGAIYCVGVDPFNETOLARIADSKDHVPVNDGFQALOGIIHSLKKSCT 221
Db      181 E 181
      222 E 222

RESULT 4
US-10-038-307-2
; Sequence 2, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
```

```
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/038.307
/ CURRENT FILING DATE: 2002-06-28
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 333
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-038-307-2
```

```
Query Match      100.0%; Score 943; DB 14; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.6e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101
QY 61 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 120
DB 102 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 161
QY 121 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 180
DB 162 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 221
QY 181 E 181
DB 222 E 222
```

```
RESULT 5
US-10-201-292-2
/ Sequence 2, Application US/10201292
/ Publication No. US20030144193A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/201.292
/ CURRENT FILING DATE: 2003-02-14
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 333
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-201-292-2
```

```
Query Match      100.0%; Score 943; DB 14; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.6e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
DB 42 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101
QY 61 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 120
DB 102 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 161
QY 121 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 180
DB 162 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 221
```

```
QY 181 E 181
DB 222 E 222
```

```
RESULT 6
US-10-038-307-22
/ Sequence 22, Application US/10038307
/ Publication No. US20030134786A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/038.307
/ CURRENT FILING DATE: 2002-06-28
/ NUMBER OF SEQ ID NOS: 26
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 342
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-038-307-22
```

```
Query Match      100.0%; Score 943; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.7e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
DB 51 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 110
QY 61 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 120
DB 111 IROGLEBLQVLPFGDITYMEHGEFRASEQIYENRQCYRTASVIALTDGELHEDLFFYS 170
QY 121 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 180
DB 171 EREANRSRDIGAIYVCVGDVFNFTOLARIADSKDHVPVNDGFOALOGIHSILKSCSI 230
QY 181 E 181
DB 231 E 231
```

```
RESULT 7
US-10-201-292-22
/ Sequence 22, Application US/10201292
/ Publication No. US20030144193A1
/ GENERAL INFORMATION:
/ APPLICANT: James B. ROTTMAN
/ APPLICANT: Theresa L. O'KEEFE
/ APPLICANT: Engin OZKAYNAK
/ APPLICANT: Judith J. HEALEY
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
/ FILE REFERENCE: 7853-253-999
/ CURRENT APPLICATION NUMBER: US/10/201.292
/ CURRENT FILING DATE: 2003-02-14
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 22
/ LENGTH: 342
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-201-292-22
```

```
Query Match      100.0%; Score 943; DB 14; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.7e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GFDLYFLDKSGSVLHHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
```



US-10-201-292-28  
/ Sequence 28, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 28  
/ LENGTH: 460  
/ TYPE: PRN  
/ ORGANISM: Homo sapiens  
US-10-201-292-28

Query Match 100.0%; Score 943; DB 14; Length 460;  
Best Local Similarity 100.0%; Pred. No. 2,5e-90;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101  
QY 61 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 120  
DB 102 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 161  
QY 121 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 180  
DB 162 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 221  
QY 181 E 181  
DB 222 E 222

RESULT 12  
US-10-201-292-32  
/ Sequence 32, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 32  
/ LENGTH: 479  
/ TYPE: PRN  
/ ORGANISM: Homo sapiens  
US-10-201-292-32

Query Match 100.0%; Score 943; DB 14; Length 479;  
Best Local Similarity 100.0%; Pred. No. 2,7e-90;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101  
QY 61 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 120  
DB 102 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 161

QY 121 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 180  
DB 162 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 221  
QY 181 E 181  
DB 222 E 222

RESULT 13  
US-10-201-292-34  
/ Sequence 34, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 34  
/ LENGTH: 504  
/ TYPE: PRN  
/ ORGANISM: Homo sapiens  
US-10-201-292-34

Query Match 100.0%; Score 943; DB 14; Length 504;  
Best Local Similarity 100.0%; Pred. No. 2,9e-90;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60  
DB 42 GFDLYFLDKSGSVLHMHNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101  
QY 61 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 120  
DB 102 IROGLEBLOKVLPGGDTYMHGFERASEQIYENRQGRASVIALTDGELHEDLFFYS 161  
QY 121 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 180  
DB 162 EREANRSRDLGAIYCVGVDFNETOLARIADSKOHVPVNDGFOALOGIHSILKKSICI 221  
QY 181 E 181  
DB 222 E 222

RESULT 14  
US-10-201-292-36  
/ Sequence 36, Application US/10201292  
/ Publication No. US20030144193A1  
/ GENERAL INFORMATION:  
/ APPLICANT: James B. ROTTMAN  
/ APPLICANT: Theresa L. O'KEEFE  
/ APPLICANT: Engin OZKAYNAK  
/ APPLICANT: Judith J. HEALEY  
/ TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
/ FILE REFERENCE: 7853-253-999  
/ CURRENT APPLICATION NUMBER: US/10/201,292  
/ CURRENT FILING DATE: 2003-02-14  
/ NUMBER OF SEQ ID NOS: 36  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 36  
/ LENGTH: 529  
/ TYPE: PRN  
/ ORGANISM: Homo sapiens  
US-10-201-292-36

```
Query Match      100.0%; Score 943; DB 14; Length 529;
Best Local Similarity 100.0%; Pred. No. 3.1e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFDLYFILDKSGSVLHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
DB      42 GFDLYFILDKSGSVLHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101
QY      61 IRQGLEELQVLPQGDITYMEHGFERASEQIYENRQGYRTASVIALTDGELHEDLFFYS 120
DB      102 IRQGLEELQVLPQGDITYMEHGFERASEQIYENRQGYRTASVIALTDGELHEDLFFYS 161
QY      121 EREANRSRDIAGAIYVCVGDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSCT 180
DB      162 EREANRSRDIAGAIYVCVGDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSCT 221
QY      181 E 181
DB      222 E 222

RESULT 15
US-10-038-307-18
; Sequence 18; Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-038-307-18

Query Match      100.0%; Score 943; DB 14; Length 551;
Best Local Similarity 100.0%; Pred. No. 3.3e-90;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GFDLYFILDKSGSVLHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 60
DB      42 GFDLYFILDKSGSVLHMNEIYFVEQLAHKFISPOLRMSFIVSTRGTTLMKLTEDREQ 101
QY      61 IRQGLEELQVLPQGDITYMEHGFERASEQIYENRQGYRTASVIALTDGELHEDLFFYS 120
DB      102 IRQGLEELQVLPQGDITYMEHGFERASEQIYENRQGYRTASVIALTDGELHEDLFFYS 161
QY      121 EREANRSRDIAGAIYVCVGDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSCT 180
DB      162 EREANRSRDIAGAIYVCVGDPNETOLARIADSKDHVPVNDGFOALOGIHSILKKSCT 221
QY      181 E 181
DB      222 E 222

Search completed: June 13, 2005, 20:36:52
Job time : 59.0354 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 13, 2005, 19:47:08 ; Search time 12.9858 Seconds  
(without alignments)  
1341.095 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_42\_222  
Perfect score: 943  
Sequence: 1 GPDLYFLDKSGSVLHNMNE.....DGFQALQGIHSLKSCIE 181

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:.\*  
1: p1r1:.\*  
2: p1r2:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	136	14.4	724	2 A48569	antigen Eml00 - B1
2	132.5	14.1	1153	1 RWHU1B	cell surface glyco
3	128	13.6	3124	2 A40020	collagen alpha 1(X
4	127.5	13.5	712	4 A45638	immunodominant mic
5	125.5	13.3	1163	1 RWHU1C	cell surface glyco
6	124.5	13.2	1153	2 S00551	leukocyte surface
7	116	12.3	3051	2 S42373	hypothetical prote
8	114.5	12.1	760	1 C2MS	classical-compleme
9	113	12.0	1747	2 A45974	collagen alpha 1(X
10	113	12.0	1857	2 S31212	collagen alpha 1(X
11	113	12.0	1888	2 S78476	collagen alpha 1(X
12	111.5	11.8	929	2 I51027	type XII collagen
13	110	11.7	2944	2 A54849	collagen alpha 1(V
14	105.5	11.2	764	1 BBHU	complement factor
15	105.5	11.2	1029	1 S21369	collagen alpha 2(V
16	103.5	11.0	567	2 T23087	hypothetical prote
17	100	10.6	460	2 T23087	hypothetical prote
18	99	10.5	698	2 B96958	dnak protein (heat
19	97.5	10.3	486	2 A37979	cartilage matrix p
20	97.5	10.3	761	1 BBMS	complement factor
21	96.5	10.2	493	2 A33809	cartilage matrix p
22	95.5	10.1	500	2 S66522	cartilage matrix p
23	95.5	10.1	1179	2 A53213	integrin alpha-E c
24	95	10.1	917	2 S09646	collagen alpha 2(V
25	95	10.1	1018	1 CGHU2A	collagen alpha 2(V
26	94	10.0	741	2 T46488	hypothetical prote
27	94	10.0	3176	2 CGHU3A	collagen alpha 3(V
28	93.5	9.9	574	2 A46283	sporozoite surface
29	93	9.9	340	2 E70121	hypothetical prote

30	92.5	9.8	559	2 S04531	thrombospondin-rel
31	92.5	9.8	843	2 A40370	undulin 1 - human
32	92.5	9.8	317	2 A37797	collagen alpha 3(V
33	88.5	9.4	932	2 JC5953	inter-alpha-inhibi
34	88	9.3	1218	2 T30889	valine-CRNA ligase
35	87.5	9.3	341	2 T32949	hypothetical prote
36	87.5	9.3	418	2 AG2350	hypothetical prote
37	87	9.2	710	2 AG2285	hypothetical prote
38	85.5	9.1	340	2 B83262	hypothetical prote
39	84.5	9.0	689	2 F84811	probable retroelem
40	84.5	9.0	800	2 AB1079	endoglucanase homo
41	84.5	9.0	875	1 ITBCAP	DNA topoisomerase
42	84.5	9.0	875	2 B91018	DNA gyrase subunit
43	84.5	9.0	875	2 D85862	DNA gyrase, subunit
44	84	8.9	181	2 B86152	hypothetical prote
45	84	8.9	487	1 VZEBPT	sensor kinase phoQ

ALIGNMENTS

RESULT 1

A48569  
antigen Eml00 - Eimeria maxima  
C/Species: Eimeria maxima  
C/Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A48569  
R:Pasamontes, L.; Hug, D.; Humbelin, M.; Weber, G.  
MOL Biochem. Parasitol. 57, 171-174, 1993  
A/Title: Sequence of a major Eimeria maxima antigen homologous to the Eimeria tenella mic  
A/Reference number: A48569; MID:93149203; PMID:842611  
A/Accession: A48569  
A/Status: Preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1724 <PAS>  
A/Cross-references: UNIPROT:Q04588; GB:M99058; NID:g158890; PID:g158891  
A/Note: Sequence extracted from NCBI backbone (NCBIN:123776, NCBI:P:123777)  
F/45-218/Domain: von Willebrand factor type A repeat homology <VWA2>  
F/238-296/Domain: thrombospondin type 1 repeat homology <THR1>  
F/309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
F/372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
F/433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
F/494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
F/560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 14.4% Score 136; DB 2; Length 724;  
Best Local Similarity 27.0%; Pred. No. 0.00074; Indels 28; Gaps 12;  
Matches 53; Conservative 36; Mismatches 79;

Qy	3	DLVFIIDKSGSV-LHHNNEIYFVEQLAHKF-ISP-QLRMSFTVFSRTGLTKLTEDRE	59
Db	47	DWLVLVDESSIGTSYNGKRSFTSNPAGTMRPLSPDVRGVLTFGSVAITRMDLSBSRA	106
Qy	60	QIRQGLEELQKLP--GGDTYMEGEFERASEQIYENRQGYR--TASVITATIDGELHED	115
Db	107	QNDLPLAAAKKLPYAAAGSTYTHGLAKA-EELIFSPKGGKRDNAPEMILMTDGA----	161
Qy	116	LFYYSERE-----AKRSRLGAIYCVGV-KQPNFQLARIA--DSKHIV-PP--VNDG	163
Db	162	-----SSRRSQTLAAEGLRRRGVILVGLGVNSAECRSIAGCDTSDTVECPYIQSN	217
Qy	164	FQALQGIHSLKSC 179	
Db	218	WGVSSQINGIIRKAC 233	

RESULT 2

RWHU1B  
cell surface glycoprotein CD11b precursor [validated] - human  
N/Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac  
eukocyte integrin alpha chain; neutrophil adherence receptor alpha chain  
C/Species: Homo sapiens (man)  
C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004

C/Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567  
 R/Corbi, A.L.; Kishimoto, T.K.; Miller, U.J.; Springer, T.A.  
 J. Biol. Chem. 263, 12403-12411, 1988  
 A/Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11B)  
 A/Reference number: A31108; MUID:88315033; PMID:2457584  
 A/Accession: A31108  
 A/Molecule type: mRNA  
 A/Residues: 1-1153 <COR>  
 A/Cross-references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148  
 A/Note: part of this sequence was confirmed by protein sequencing  
 R/Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
 J. Cell Biol. 106, 2153-2158, 1988  
 A/Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M  
 A/Reference number: A28915; MUID:88257215; PMID:2454931  
 A/Accession: A28915  
 A/Molecule type: mRNA  
 A/Residues: 1-499, 501-965, 'P', 967-1153 <ARN>  
 A/Cross-references: GB:M18044; GB:J03370; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594  
 A/Note: the authors translated the codon TAC for residue 1129 as Thr  
 R/Shelley, C.S.; Arnaout, M.A.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991  
 A/Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
 A/Reference number: A41600; MUID:92073318; PMID:1683702  
 A/Accession: A41600  
 A/Molecule type: DNA  
 A/Residues: 1-9 <SHE>  
 A/Cross-references: GB:M76724; NID:G180018; PIDN:AAA8410.1; PID:G553215  
 R/Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
 A/Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion  
 A/Reference number: A94193; MUID:88190151; PMID:2833753  
 A/Accession: A30892  
 A/Molecule type: mRNA  
 A/Residues: 917-1042 <AR2>  
 A/Cross-references: GB:M18044  
 R/Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
 A/Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
 A/Reference number: A32218; MUID:89098893; PMID:2563162  
 A/Accession: A32218  
 A/Molecule type: mRNA  
 A/Residues: 9-1153 <HIC>  
 A/Cross-references: GB:J04145; NID:G189068; PIDN:AAA59903.1; PID:G386975  
 A/Note: part of this sequence was confirmed by protein sequencing  
 R/Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
 J. Immunol. 150, 480-490, 1993  
 A/Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-in  
 n during evolution.  
 A/Reference number: A46526; MUID:93123748; PMID:8419480  
 A/Accession: A46526  
 A/Status: not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-499, 501-1153 <FLE>  
 A/Cross-references: GB:552227; NID:G263047; PIDN:AA24821.1; PID:G263049  
 A/Note: the last three bases of intron 13, CAG, are included in some but not all mature  
 R/Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.  
 Biochim. Biophys. Acta 874, 368-371, 1986  
 A/Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
 A/Reference number: A90664; MUID:8706671; PMID:3539202  
 A/Accession: A26091  
 A/Molecule type: protein  
 A/Residues: 17-31 <PIE>  
 A/Experimental source: granulocytes  
 R/Pahl, H.L.; Rosemarin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992  
 A/Title: Characterization of the myeloid-specific CD11b promoter.  
 A/Reference number: 152567; MUID:92144986; PMID:11346576  
 A/Accession: 152567  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: DNA

A/Residues: 1-9 <RES>  
 A/Cross-references: GB:M84477; NID:G180184; PIDN:AAA51960.1; PID:G553219  
 C/Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1 (C3)  
 C/Genetics:  
 A/Gene: GDB:ITGAM, CR3A  
 A/Cross-references: GDB:120599; OMIM:120980  
 A/Map position: 16p11.2-16p11.2  
 A/Note: promoter contains a GATA motif and two Sp1 consensus binding sites  
 C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
 C/Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; magr  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>  
 F:17-1108/Domain: extracellular #status predicted <EXT>  
 F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F:465-473/Region: calcium/magnesium binding #status predicted  
 F:530-538/Region: calcium/magnesium binding #status predicted  
 F:593-601/Region: calcium/magnesium binding #status predicted  
 F:1109-1134/Domain: transmembrane #status predicted <TM>  
 F:1135-1153/Domain: intracellular #status predicted <INT>  
 F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
 Query Match 14.1%; Score 132.5; DB 1; Length 1153;  
 Best Local Similarity 26.3%; Pred. No. 0.0026;  
 Matches 52; Conservative 41; Mismatches 66; Indels 39; Gaps 11;  
 QY 3 DLYFIIDKSGSVL-FHMEIYVFEQLAKFISPOLMSFVSTGTITMKTED---- 57  
 DB 150 DIAPLIDSGSILPHDFRMRKEFVST-----VNEOLKSKSTLPS-----LMOYSEPRRIH 199  
 QY 58 -----REQIQGLEELQKVRPGDTHMGSEFASBOIYVENQGVTA-SVIALT 108  
 DB 200 FTFKERQNNPNPGLVAPITQTL--GRHTATGIRKVRRELFNTTGARKNAKILVIT 257  
 QY 109 DGEIHRDLDFEYSB--SEANRSRDIGATVYCVGVKDFNETQLAR-----IADS--KDHVP 159  
 DB 258 DGEKFDPLGYEDVIREADRE--GVIRYIVGVGDAFRSKSKQELNTIAKSPRDHVFQ 314  
 QY 160 VNDGFQALGCIHSITLK 177  
 DB 315 VNN-PEALKTQIMQLEK 331  
 RESULT 3  
 A40020  
 collagen alpha 1(XII) chain precursor - chicken  
 N/Alternate names: fibrochmerin  
 C/Species: Gallus gallus (chicken)  
 C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 09-Jul-2004  
 C/Accession: A40020; A34485; B34485; A28037; S23814; S22554; S28811  
 R/Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obara  
 J. Cell Biol. 115, 209-221, 1991  
 A/Title: The complete primary structure of type XII collagen shows a chimeric molecule w  
 nous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.  
 A/Reference number: A40020; MUID:92011862; PMID:1918137  
 A/Accession: A40020  
 A/Molecule type: mRNA  
 A/Residues: 1-3124 <YAM>  
 A/Cross-references: UNIPROT:P13944; GB:D00824; NID:G222810; PIDN:BA400701.1; PID:G222811  
 A/Note: in the authors' translation residues 1216-1219 are shown after residue 1235 and,  
 R/Gordon, M.K.; Gerecke, D.R.; Duble, B.; van der Reet, M.; Olsen, B.R.  
 J. Biol. Chem. 264, 19772-19778, 1989  
 A/Title: Type XII collagen. A large multidomain molecule with partial homology to type IX  
 A/Reference number: A34485; MUID:90062079; PMID:2584192  
 A/Accession: A34485  
 A/Molecule type: mRNA  
 A/Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>  
 A/Cross-references: EMBL:U05137; NID:G211284; PIDN:AAA48635.1; PID:G211285  
 A/Accession: B34485  
 A/Molecule type: protein  
 A/Residues: 2772-2792, 2846-2873 <GOR2>  
 R/Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987  
 A/Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c]



A/Reference number: A28037; MUID:87317590; PMID:3476925  
 A/Accession: A28037  
 A/Molecule type: mRNA  
 A/Releases: 2960-2976, 'P', 2978-3074, 'AG' <GORG>  
 A/Cross-References: EMBL:M13755; NID:9211649; PIDN:AAA48718.1; PID:9211650  
 A/Note: This sequence has been revised in reference A34485  
 A/Koch, M.; Bernasconi, C.; Chiquet, M.  
 Eur. J. Biochem. 207, 847-856, 1992  
 A/Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of  
 A/Release number: 823814; MUID:92362621; PMID:1323460  
 A/Accession: 823814  
 A/Molecule type: Protein  
 A/Releases: 'X', 1333, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517  
 R/Dubler, B.; van der Rest, M.  
 J. Biol. Chem. 262, 17724-17727, 1987  
 A/Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin-  
 A/Release number: 822254; MUID:88087065; PMID:3121603  
 A/Accession: 822254  
 A/Molecule type: Protein  
 A/Releases: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DUB>  
 R/Tueb, U.; Tueb, B.  
 Biochim. Biophys. Acta 1171, 97-98, 1992  
 A/Title: The two splice variants of collagen XII share a common 5' end.  
 A/Reference number: 828811; MUID:93042014; PMID:1420368  
 A/Accession: 828811  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Releases: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>  
 A/Cross-References: EMBL:X67327  
 C/Genetics:  
 A/Intons: 2845/3, 2863/3, 2887/3, 2899/3, 2922/1, 2985/1, 3008/1, 3065/1  
 C/Keywords: alternative splicing; cell binding; coiled coil; connective tissue; disulfid  
 F/1-23/Domains: signal sequence #status predicted <SIG>  
 F/24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>  
 F/24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
 F/24-114/Domains: fibronectin type III repeat homology <FNI>  
 F/24-105/Domains: fibronectin type III repeat homology <FNA>  
 F/13-301/Domains: von Willebrand factor type A repeat homology <VWA1>  
 F/33-425/Domains: IIB #status predicted <IIB>  
 F/33-414/Domains: fibronectin type III repeat homology <FN3B>  
 F/437-601/Domains: von Willebrand factor type A repeat homology <VWA2>  
 F/629-1178/Domains: IIC #status predicted <IIC>  
 F/630-711/Domains: fibronectin type III repeat homology <FN3C>  
 F/721-802/Domains: fibronectin type III repeat homology <FN3D>  
 F/812-895/Domains: fibronectin type III repeat homology <FN3B>  
 F/905-986/Domains: fibronectin type III repeat homology <FN3F>  
 F/1995-1076/Domains: fibronectin type III repeat homology <FN3G>  
 F/1086-1169/Domains: fibronectin type III repeat homology <FN3H>  
 F/1187-1361/Domains: von Willebrand factor type A repeat homology <VWA3>  
 F/1384-2295/Domains: IID #status predicted <IID>  
 F/1384-1465/Domains: fibronectin type III repeat homology <FN3I>  
 F/1474-1557/Domains: fibronectin type III repeat homology <FN3J>  
 F/1566-1647/Domains: fibronectin type III repeat homology <FN3K>  
 F/1565-1738/Domains: fibronectin type III repeat homology <FN3L>  
 F/1756-1838/Domains: fibronectin type III repeat homology <FN3M>  
 F/1847-1928/Domains: fibronectin type III repeat homology <FN3N>  
 F/1937-2019/Domains: fibronectin type III repeat homology <FN3O>  
 F/2028-2110/Domains: fibronectin type III repeat homology <FN3P>  
 F/2119-2199/Domains: fibronectin type III repeat homology <FN3Q>  
 F/2207-2294/Domains: fibronectin type III repeat homology <FN3R>  
 F/2335-2490/Domains: von Willebrand factor type A repeat homology <VWA4>  
 F/2438-2440/Domains: cell adhesion #status predicted  
 F/2509-2750/Domains: IXP, homologous to NC4 domain of type IX collagen #status predicted  
 F/2751-2902/Domains: collagenous CO2 #status predicted <COL2>  
 F/2889-2901/Domains: cell attachment (R-G-D) motif  
 F/2903-2945/Domains: non-collagenous NC2 #status predicted <NC2>  
 F/2946-3048/Domains: collagenous COL1 #status predicted <COL1>  
 F/3049-3124/Domains: non-collagenous NC1 #status predicted <NC1>  
 F/3127, 1006, 1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carboxyproline (Asn)  
 P/2780, 2789, 2836, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #stat  
 Query Match 13.6%; Score 128; DB 2; Length 3124;  
 Best Local Similarity 26.3%; Pred. NO. 0.021;

	Matches	49; Conservative	42; Mismatches	79; Indels	16; Gaps	9;
Qy	3	DLVPIILDKSGSV-LAHNNELIYFVEQLAHK--ISQP-LRMSFVFSRGTLMKLT--E	56			
Db	2327	DIVLTALTAWSIGDNNKVKVFFVNTVGADLINPAGIQVSLVQSDBAOSEFKLNTD	2386			
Qy	57	DREQIRQGLBELQKVLPGCDTYMHEGEPASSEQIY-YENRQSVRTASVIALITDGEIHLAD	115			
Db	2387	DKAALAGALGNVQ--YRGAMTRRGKALTPIKEKYLTMESGMRGRVPKLVVLTGDRSGD	2444			
Qy	116	LEFYSEREARNRDLGAIIVYCVGKDFNFTQLAIAD--SKDHYFPVNDGQALQGIHS	173			
Db	2445	V---RKAATVICHSGFSVFVGVGADVDVYNELAKIASPSERHVFIYDD--FDAEKIQDN	2499			
Qy	174	ILKKSC	179			
Db	2500	LVTFCV	2505			

[illegible]

```

A;Reference number: A36584
A;Contents: extractum
A;Accession: A36584
A;Molecule type: DNA
A;Residues: 1-1163 <COR>
A;Cross-references: UNIPROT:P20702
A;Note: this revision to the sequence from reference A35543 includes the carboxyl end
R;Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 2782-2788, 1990
A;Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.
A;Reference number: A35543; MUID:90153906; PMID:2303426
A;Accession: A35543
A;Molecule type: DNA
A;Residues: 1-834 <CO2>
A;Note: this sequence has been revised in reference A36584
R;Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.
EMBO J. 6, 4023-4028, 1987
A;Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte
A;Reference number: S00864; MUID:88166645; PMID:3327687
A;Accession: S00864
A;Molecule type: mRNA
A;Residues: 1-755; 'L', 757-1163 <CO3>
A;Cross-references: GB:M61695; EMBL:Y00093; NID:9487829; PIDN:AAA59180.1; PID:9487830
A;Note: part of this sequence was confirmed by protein sequencing
C;Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my
C;Genetic8:
A;Gene: GDB:ITGAX; CD11C
A;Cross-references: GDB:119758; OMIM:151510
A;Map position: 16p11.2-16p11.2
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C;Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>
F;149-319/Domain: extracellular #status predicted <EXT>
F;1108-1133/Domain: von Willebrand factor type A repeat homology <VMA4>
F;1134-1163/Domain: transmembrane #status predicted <TM>
F;61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre
Query Match 13.3%; Score 125.5; DB 1; Length 1163;
Best Local Similarity 24.7%; Pred. No. 0.01;
Matches 48; Conservative 38; Mismatches 77; Indels 31; Gaps 9;
QY 3 DLYFLIDKSGSV-LAHNNEIYFVEQLAHKFI-SQLRMSFIVSTGTTLMLKLTEDR-- 59
DB 151 DIVFLIDGSGSISRNFAVMNFRAVISQFQSTQFSLMQFNKQFHTFEERFRTS 210
QY 60 ---QIRGLEBELQKVLPGDPTVMHEGFERASEQIYVENRQGYRTAS-VITALTDELHED 115
DB 211 NPISLASVAVQLQ-----GFTYTAIAQNVVHRLPHASYGARRKATKILVIITDGKKEGD 265
QY 116 LFFYSEREANRSRDGAIVYCVGV-----KDNETQLARIAD--SKDHVPVNDG 163
DB 266 SLVDKQD-VIEMADAGIIRYAVIGLAFQNRNRSKEIND-----TASKPQEHIFKVED- 318
QY 164 FOALQGIHSLKK 177
DB 319 FDALKDIONQLKEX 332

```

## RESULT 6

```

S00551
Leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse
N;Alternate names: complement-3 receptor alpha chain
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
A;Accession: S00551; I59078
R;Pyrcala, R.
EMBO J. 7, 1371-1378, 1988
A;Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the
A;Reference number: S00551; MUID:88312584; PMID:3044779
A;Accession: S00551
A;Molecule type: DNA

```

```

A;Residues: 1-1153 <PYT>
A;Cross-references: UNIPROT:P05555; EMBL:X07640; NID:952982; PIDN:CAA30479.1; PID:952983
A;Note: the authors translated the codon CAC for residue 569 as Gln
R;Sastre, L.; Roman, J.M.; Teplov, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,
Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986
A;Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recep
A;Reference number: I59078; MUID:86287312; PMID:2942940
A;Accession: I59078
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 11-44 <RES>
A;Cross-references: GB:M14293; NID:g198993; PIDN:AAA39484.1; PID:g554193
C;Genetic8:
A;Gene: Mac-1
C;Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C;Keywords: cell adhesion; glycoprotein; transmembrane protein
F;1-16/Domain: signal sequence #status predicted <SIG>
F;11-153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental
F;148-318/Domain: von Willebrand factor type A repeat homology <VMA2>
F;1106-1129/Domain: transmembrane #status predicted <TM>

```

Query Match 13.2%; Score 124.5; DB 2; Length 1153;

Best Local Similarity 26.8%; Pred. No. 0.013;

Matches 53; Conservative 37; Mismatches 69; Indels 39; Gaps 12;

```

QY 3 DLYFLIDKSGSV-LAHNNEIYFVEQLAHKFI-SQLRMSFIVSTGTTLMLKLTED-- 57
DB 150 DIVFLIDGSGSI-----NNIDF---QKKEFVSFTMEQFKSKTLFS-----LMQYSDPFR 197
QY 58 ---REGIRGLEBELQKVLPGDPTVMHEGFERASEQIYVE--NRQGYRTASVITALT 108
DB 198 IHTFTNDFKNSBPBRSHVSIKQLNGRTKTASGIRKVRLELPFKTNGARENAKILVIT 257
QY 109 DGEIHEDLFFYSR--REANRSRDGAIVYCVGYKD--FNETQLARIAD-----SKDHVP 159
DB 258 DGEKFGDPLDKVIREADRA---GVIRYVIGGNAPNKRQSRREDLTASKRAGEHVFQ 314
QY 160 VNDFOALQGIHSLKK 177
DB 315 V-DNFEALNTIQNLQLEK 331

```

## RESULT 7

```

S42373
hypothetical protein T20G5.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 09-May-2004
A;Accession: S42373
R;Smith, A.
Submitted to the EMBL Data Library, March 1994
A;Reference number: S42368
A;Accession: S42373
A;Molecule type: DNA
A;Residues: 1-3051 <SMI>
A;Cross-references: EMBL:Z30423; NID:g458479; PID:g458485
C;Genetic8:
A;Intons: 96/1; 166/3; 201/2; 355/1; 404/1; 427/2; 794/1; 892/1; 952/2; 1041/1; 1146/1;
F;512-679/Domain: von Willebrand factor type A repeat homology <VMA1>
F;754-793/Domain: fibronectin type II repeat homology <271>
F;1201-1244/Domain: EGF homology <EGF>

```

Query Match 12.3%; Score 116; DB 2; Length 3051;

Best Local Similarity 28.2%; Pred. No. 0.22;

Matches 55; Conservative 32; Mismatches 74; Indels 34; Gaps 11;

```

QY 3 DLYFLIDKSGSV-LAHNNEIYFVEQLAHKFI-SQLRMSFIVSTGTTLMLKLTED-- 56
DB 514 DLVFLIDGSGSISYFKNEVLRVREYVLEFISGKTRVGILQYSDQIRHEPDLQYG 573
QY 57 DREQIRGLEBELQKVLPG-----GDTVMH---EGFERASEQIYVENRQGYR-----TASVI 104
DB 574 DRDSLKIGSIFQ-VLTGLTRGALQIHVQSGF-----SEKRGARPGQSDIARVA 623

```

Oy 105 IALTDLGHELDLFFYSREANRSRDLCALIVCVGVKQ-FNETQLARIADSKDHVPVNDG 163  
Db 624 IITLTDGSDQNV-----TGPAIDSAKRLSINTPAIGVTDHVLASLESIASFNRFVY-DK 678  
Oy 164 FOALOGIHSILKKS 178  
Db 679 FKDLDTRLRSWIKQA 693

## RESULT 8

C2MS

N/Alternate names: C3 convertase; C5 convertase; complement C2  
C/Species: Mus musculus (house mouse)  
C/Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text\_change 09-Jul-2004  
C/Accession: A38876; B36593; I54429  
R/Ishtawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
Submitted to GenBank, January 1991  
A/Reference number: A38875  
A/Accession: A38876  
A/Molecule type: DNA  
A/Residues: 1-760 <I82>  
A/Cross-references: UNIPROT:P21180; GB:M57891; GB:J05661; NID:g192436; PIDN:AAA63294.1;  
R/Ishtawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
J. Biol. Chem. 265, 19040-19046, 1990  
A/Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different me  
A/Reference number: A36593; MUID:91035430; PMID:2229060  
A/Accession: B36593  
A/Molecule type: mRNA  
A/Residues: 1-760 <I8>  
A/Cross-references: EMBL:M57891; NID:g192436; PIDN:AAA63294.1; PID:g192437  
R/Palu, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, J.; Whitehead, A.S.; Colten, H.R.  
Immunogenetics 25, 290-298, 1987  
A/Title: DNA polymorphism of H2C III genes in inbred and wild mouse strains.  
A/Reference number: I54429; MUID:87192938; PMID:2883115  
A/Accession: I54429  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 660-677, 'R', 679, 681-723, 'G', 725 <RES>  
A/Cross-references: GB:M16271; NID:g199289; PIDN:AAA39562.1; PID:g199290  
C/Genetic: 16/1, 91/1, 153/1, 212/1, 245/3, 290/3, 337/1, 384/1, 414/1, 461/1, 492/3, 51  
A/Introns: 16/1, 91/1, 153/1, 212/1, 245/3, 290/3, 337/1, 384/1, 414/1, 461/1, 492/3, 51  
C/Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C2a,  
C/Function:  
A/Description: cleaves complement C3 and complement C5 alpha chains  
A/Pathway: complement classical pathway  
C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; v  
C/Keywords: alternative splicing; complement classical pathway; duplication; glycoprotein  
F/1-18/Domain: signal sequence #status predicted <SIG>  
F/19-250/Product: complement C2b fragment #status predicted <C2b>  
F/12-69/Domain: complement factor H repeat homology <FH1>  
F/94-149/Domain: complement factor H repeat homology <FH2>  
F/156-210/Domain: complement factor H repeat homology <FH3>  
F/51-160/Product: complement C2a fragment long form #status predicted <C2a>  
F/51-605, 613-760/Product: complement C2a fragment short form #status predicted <C2s>  
F/559-449/Domain: von Willebrand factor type A repeat homology <VFA>  
F/478-747/Domain: trypsin homology #status atypical <TRY>  
F/12-69, 48-89, 94-136, 122-149, 156-197, 182-210, 470-550, 499-515, 593-609, 647-674, 685-715/D18  
F/17, 117, 297, 340, 474, 478, 663/Binding site: carbohydrate (Asn) (covalent) #status predict  
F/514, 570, 689/Active site: His, Asp, Ser #status predicted

Query Match 12.1%; Score 114.5; DB 1; Length 760;  
Best Local Similarity 24.3%; Pred. No. 0.056; Matches 53; Mismatches 72; Indels 49; Gaps 10;

Oy 3 DLVFIIDKSGSVLHHNNEIY-VFVEQLAHKFIISPOL--MSFVIFSTRGTIMKLTRE 59  
Db 261 NLVLLDASQSVTEKFDIFKSAELMVERIPSEFVAVVAITFAQPKTMSILSERS 320  
Oy 60 QIHGGLBELOKVLPGDPTVHHEGFERSAQIYVE-----NRGYSRTAS---- 102  
Db 321 Q-----DVEVITSDSASVSKDHENATYGVLLIRVYSPMOTOMRDLGMEISAWKEI 374

Oy 103 -VITALDGLHEDLFFYSREANRSRDLCAL-----VTCGVY-----KQFNE 144  
Db 375 RHITLITLTDGK--SNMGDSPKKAVTRIRIELISIQNRNDYDIYAIGVKDLDVWKEINE 432  
Oy 145 TQLARIADSKDHVPVNDGFOALOGIHSILKKS 178  
Db 433 --LGSKKDGERHAFILQDA-KALQOIFEHMLDVS 463

## RESULT 9

A45974

collagen alpha 1(XIV) chain precursor, short form 2 - chicken  
N/Alternate names: undulin  
C/Species: Gallus gallus (chicken)  
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: A45974; S30085; S22916; S17035; S20833  
R/Gerecke, D.R.; Foley, J.W.; Castagnola, P.; Gennari, M.; Dublet, B.; Cancedda, R.; Lin  
J. Biol. Chem. 268, 12177-12184, 1993  
A/Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' region  
ns.  
A/Reference number: A45974; MUID:93280195; PMID:8505337  
A/Accession: A45974  
A/Status: preliminary  
A/Molecule type: mRNA; protein  
A/Residues: 1-1747 <GER>  
A/Cross-references: UNIPROT:P32018  
A/Experimental source: embryo skin  
A/Note: sequence inconsistent with the nucleotide translation  
A/Note: sequence extracted from NCBI backbone (NCBIN:133364, NCBI:P.133365)  
R/Apte, S.S.  
Submitted to the EMBL Data Library, March 1992  
A/Reference number: S30085  
A/Accession: S30085  
A/Molecule type: mRNA  
A/Residues: 1472-1660 <APT>  
A/Cross-references: EMBL:X65122; NID:962871; PIDN:CAA46238.1; PID:9938175  
R/Trueb, J.; Trueb, B.  
Eur. J. Biochem. 207, 549-557, 1992  
A/Title: Type XIV collagen is a variant of undulin.  
A/Reference number: S22916; MUID:92339443; PMID:1339349  
A/Accession: S22916  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 286-494, 'Q', 496-834, 'A', 836-1119, 'KL', 1122-1402, 1409-1439 <TRU>  
R/Gordon, M.K.; Castagnola, P.; Dublet, B.; Linsemayer, T.F.; van der Ree, M.; Mayne, R.  
Eur. J. Biochem. 201, 333-338, 1991  
A/Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens  
A/Reference number: S17035; MUID:92037585; PMID:1935930  
A/Accession: S17035  
A/Molecule type: mRNA  
A/Residues: 1472-1659 <GOR1>  
A/Accession: S20833  
A/Molecule type: protein  
A/Residues: 1551-1570, 1593-1599, 1639-1667 <GOR2>  
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer  
F/40-204/Domain: von Willebrand factor type A repeat homology <VWA1>  
F/326-409/Domain: fibronectin type III repeat homology <FN3A>  
F/326-409/Domain: fibronectin type III repeat homology <FN3B>  
F/418-498/Domain: fibronectin type III repeat homology <FN3C>  
F/507-591/Domain: fibronectin type III repeat homology <FN3D>  
F/625-707/Domain: fibronectin type III repeat homology <FN3E>  
F/716-798/Domain: fibronectin type III repeat homology <FN3F>  
F/806-893/Domain: fibronectin type III repeat homology <FN3G>  
F/924-1089/Domain: fibronectin type A repeat homology <FN3H>  
F/111-135/Domain: non-collagenous NC4 #status predicted <NC4>  
F/111-155/Domain: non-collagenous NC2 #status predicted <NC2>  
F/1554-1659/Domain: triple helical domain COL1 #status predicted <COL1>

Query Match 12.0%; Score 113; DB 2; Length 1747;  
Best Local Similarity 26.1%; Pred. No. 0.2; Matches 49; Mismatches 39; Indels 20; Gaps 11;

Oy 3 DLVFIIDKSGSV-LHHNNEIYFVEQL--AHKFIISPOL-LRMSFVIFSTRGTIMKLT--E 56



	Query Match	11.8%; Score 111.5; DB 2; Length 929;
	Best Local Similarity	23.2%; Pred. No. 0.13; Mismatches 43; Conservative 37; Indels 15; Gaps 6;
Oy	3 DLYFIIDKSSV--LHHNNEIYFVEQLAHKFISPOLRMSFLVFSTRTGTTIMKLTEDREQI 61	
Db	633 DIVLLLVGSGWSIGHPNFKIYVNFISRVAEVEDISDDEVOIALVSQSDPRTMOLNTHTKT 692	
Oy	62 ROGLEELDKVLPR--GGPTVMHGEFERASEQIYYENRGY----RTASVITALTDELGHED 115	
Db	693 KKSILMDAVANLLPYKGSGNTNGSALKLFLENPF---RFVGCKREARAKIALILLTDKSKDD 749	
Oy	116 LFPYSEREAANSRDGLAIVCVGVKDFNETQTALIAISKDHVPFN--DGFOALQGIIHSI 174	
Db	750 IVAPSKRYA----DEGIELVAYGIKNADENLETIASDPDELYMYNVADPSLLTNIVDL 805	
Oy	175 LKKSC 179	
Db	806 TENVC 810	

### RESULT 13

collagen alpha 1(VII) chain precursor - human  
N/Alternate names: procollagen alpha 1 (VII) chain  
C/Species: Homo sapiens (man)  
C/Date: 04-Nov-1994 #sequence\_revision 04-Nov-1994 #text\_change 09-Jul-2004  
A/Accession: A54849, PH0844, S16316, I56328, A30296, I84686  
R/Christiano, A.M.; Greenpan, D.S.; Lee, S.; Uitto, J.  
J. Biol. Chem. 269, 20256-20262, 1994  
A/Title: Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII)  
A/Reference number: A54849, MUID:94327588, PMID:8051117  
A/Accession: A54849  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 1-2944 <CHR>  
A/Cross-references: UNIPROT:Q02388; GB:I02870; NID:G987124; PIDN:AAV5438.1; PID:G987125  
R/Tenaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A/Title: Molecule cloning and characterization of type VII collagen cDNA.  
A/Reference number: PH0844; MUID:92231902; PMID:1567409  
A/Accession: PH0844  
A/Molecule type: mRNA  
A/Residues: 1-2944  
A/Cross-references: DDBJ:D11152; DDBJ:D13694; NID:G453698; PIDN:BA02853.1; PID:G453699  
A/Experimental source: Keratinocyte  
A/Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
R/Parente, M.G.; Chung, L.C.; Rymaszewski, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991  
A/Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A/Reference number: S16316; MUID:91334380; PMID:1871109  
A/Accession: S16316  
A/Molecule type: mRNA  
A/Residues: 815-892, 'E', 894-1439 <PAR>  
A/Cross-references: GB:M5618; GB:S49017; NID:G180914; PIDN:AAA96439.1; PID:G180915  
A/Experimental source: Keratinocyte  
R/Gemmon, W.R.; Abernethy, M.L.; Padilla, K.M.; Pilsbayan, P.S.; Cook, M.E.; Wright, J.; U  
J. Invest. Dermatol. 99, 691-696, 1992  
A/Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protob  
A/Reference number: I56328; MUID:93107742; PMID:1469284  
A/Accession: I56328  
A/Status: translated from GB/EMBL/DDBJ  
A/Molecule type: mRNA  
A/Residues: 1-2944  
A/Cross-references: GB:S51236; NID:G262308; PIDN:AAH24537.1; PID:G262309  
R/Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, R.W.; Burgess, R.E.  
J. Biol. Chem. 264, 3822-3826, 1989  
A/Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
A/Reference number: A30286; MUID:89139437; PMID:2537292  
A/Accession: A30286  
A/Molecule type: Protein  
A/Residues: 'A', 1240-1246, 'G', 1248-1250, 'X', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, 'V'  
A/Note: two reported peptides cannot be reliably located

R:Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A>Title: The carboxyl-terminal half of type VI collagen, including the non-collagenous ;  
A:Reference number: 148103; MUID:93271985; PMID:849916  
A:Accession: 184686  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 2395-2871, 'S', 2873-2944 <R2>  
A:Cross-references: GB:106862; NID:9388713; PID:AAA69196.1; PID:9388714  
R:Christiano, A.M.; RYmaenen, M.; Uitto, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A>Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly--> Ser subse  
A:Reference number: A55255; MUID:94224777; PMID:8170945  
A:Contents: annotation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
ed and subsequently O-glycosylated.

A:Gene: GDB:COL1; EBR1; EBD1; EB  
A:Cross-references: GDB:128750; OMIM:120120  
A:Map position: 3p21.3-3p21.3  
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis  
A:Note: there are 118 introns  
C:Complex: type VII collagen is probably a homotrimer  
C:Function:  
A:Description: structural component of extracellular polymer associated with anchoring fibrils  
C:Keywords: collod coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:117-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F:117-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:316-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:323-318/Domain: fibronectin type III repeat homology <FN1>  
F:327-413/Domain: fibronectin type III repeat homology <FN2>  
F:414-502/Domain: fibronectin type III repeat homology <FN3>  
F:508-593/Domain: fibronectin type III repeat homology <FN4>  
F:598-683/Domain: fibronectin type III repeat homology <FN5>  
F:688-771/Domain: fibronectin type III repeat homology <FN6>  
F:778-862/Domain: fibronectin type III repeat homology <FN7>  
F:864-952/Domain: fibronectin type III repeat homology <FN8>  
F:954-1045/Domain: fibronectin type III repeat homology <FN9>  
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1110-1172/Region: cell attachment (R-G-D) motif  
F:1189-1253/Region: cysteine/proline-rich  
F:1254-2783/Region: interrupted helical  
F:1334-1336/Region: cell attachment (R-G-D) motif  
F:2008-3010/Region: cell attachment (R-G-D) motif  
F:2553-2555/Region: cell attachment (R-G-D) motif  
F:2784-2844/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F:2876-2929/Domain: animal knitr-type proteinase inhibitor homology <BPI>  
F:337,786,1109/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:2157,2176,2185,2188,2664,2677,2673/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:2625,2631/Binding site: carboxylate (Lys) (covalent) #status experimental  
F:2634,2802,2804/Dissulfide bonds: interchain #status predicted

Query Match	11.7%	Score 110;	DB 2;	Length 2944;
Best Local Similarity	26.2%	Pred. No. 0.69;		
Matches	49;	Conservative 37;	Mismatches 83;	Indels 18; Gaps 10;
QY	3	DLVYILDKGSGV-LHHNNEIYFVEQLAHK---ISPO-IRMSTIVSTRGIT--LMKL	54	
DB	38	DIIVLVLQSSISIGRNSFEVRSFLEGLVLPEPSGAASQGVAFVQYSDDRTEFGDAL	97	
QY	55	TEDEBQIRQGEELQKVLPPGDDTYVMHGEGFEASBOIYENRQYRTASVIALPDGELHE	114	
DB	98	GSGGDVIR-AIRELS--YKGSNTTGGAIIHVAHDVFLPOLARGVPRVCLIIDTK-SQ	153	
QY	115	DLFFYSEERANRSSDLGAIYVCVGKQNETQLARIAD--SKDHVPYNDGFQALQGITI	172	
DB	154	DLV---DTAAQRLLKGQGVLFVAVGIIKNADPEELKRVASQPTSDFFEVND--FSIIRTL	209	
QY	173	SLIKKSC	179	
DB	210	LVSRRVC	216	

## RESULT 14

BBNU  
 Complement factor B precursor (validated) - human  
 N:Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-labile; alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment  
 C:Species: Homo sapiens (man)  
 C>Date: 19-Feb-1984 #sequence, revision 05-Aug-1994 #text change 09-Jul-2004  
 C/Accession: S34075; A44622; A00934; A19188; A19947; B19947; B25971; S14339; A44628; 154  
 R:Mejla, J.E.; Jahn, I.; de la Salle, H.; Hauptmann, G.  
 Submitted to the EMBL Data Library, March 1993  
 A:Reference number: S34075  
 A/Accession: S34075  
 A/Molecule type: mRNA  
 A/Residues: 1-764 <MBJ>  
 A/Cross-references: UNIPROT:P00751; EMBL:X72875; NID:g297568; PIDN:CAA51389.1; PID:g297568  
 R:Moore, D.E.; Markham, A.F.; Ricker, A.T.; Goldberger, G.; Colten, H.R.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
 A>Title: Isolation of cDNA clones for the human complement protein factor B, a class III  
 A/Reference number: A44622; MUID:83039428; PMID:6957884  
 A/Accession: A44622  
 A/Molecule type: mRNA  
 A/Residues: 467-546; 550-595; 752-764 <WOO>  
 A/Cross-references: GB:J00185; GB:J00186  
 A>Note: the authors translated the codon TAC at 519 as Thr; the nucleic acid translation  
 R:Moyle, J.E.; Anderson, J.K.; Davison, E.A.; Woods, D.E.  
 J. Biol. Chem. 259, 3407-3412, 1984  
 A>Title: Complete primary structure for the zymogen of human complement factor B.  
 A/Reference number: A20751; MUID:84161997; PMID:6546754  
 A/Accession: A00934  
 A/Molecule type: protein; mRNA  
 A/Residues: 26-764 <MOJ>  
 A/Cross-references: GB:K01566  
 A>Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
 A>Note: glycosylation sites were determined  
 R:Christie, D.L.; Gagnon, J.  
 Biochem. J. 209, 61-70, 1983  
 A>Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th  
 A/Reference number: A19188; MUID:83204002; PMID:6342610  
 A/Accession: A19188  
 A/Molecule type: protein  
 A/Residues: 260-296; T, 298-764 <CHR>  
 R:Campbell, R.D.; Porter, R.R.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
 A>Title: Molecular cloning and characterization of the gene coding for human complement  
 A/Reference number: A19947; MUID:83273641; PMID:6308626  
 A/Accession: A19947  
 A/Molecule type: DNA  
 A/Residues: 346-764 <CAM>  
 A/Cross-references: GB:U00125  
 A/Accession: B19947  
 A/Molecule type: mRNA  
 A/Residues: 339-509 <CAI>  
 A/Cross-references: GB:U00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
 R:Wu, L.; Morley, B.J.; Campbell, R.D.  
 Cell 48, 331-342, 1987  
 A>Title: Cell-specific expression of the human complement protein factor B gene: evidenc  
 A/Reference number: A25971; MUID:87102880; PMID:3643061  
 A/Accession: B25971  
 A/Molecule type: DNA  
 A/Residues: 1-99 <WU>  
 A/Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
 R:Niemann, M.A.; Bhown, A.S.; Miller, E.J.  
 Biochem. J. 274, 473-480, 1991  
 A>Title: The principal site of glycation of human complement Factor B.  
 A/Accession: S14339  
 A/Molecule type: protein  
 A/Residues: 270-329 <NIR>  
 A>Note: binding site for carbohydrate to lysine under artificial conditions

R:Morley, B.J.; Campbell, R.D.  
 EMBL J. 3, 153-157, 1984  
 A>Title: Internal homologies of the Ba fragment from human complement component factor B,  
 A/Reference number: A44628; MUID:84158524; PMID:6323161  
 A/Accession: A44628  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 16-225; F, 227-259 <MOR>  
 R:Schwabe, W.; Lutting, B.; Sokolowski, T.; Betscher, C.; Weiss, E.H.; Meyer zum Buscher  
 Immunobiology 188, 221-232, 1993  
 A>Title: Human complement factor B: functional properties of a recombinant zymogen of the  
 A/Reference number: 154409; MUID:94041399; PMID:8225386  
 A/Accession: 154409  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-764 <RBS>  
 A/Cross-references: GB:S67310; NID:g452937; PIDN:AAD13989.1; PID:g4261689  
 R:Horluchi, T.; Kim, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
 Mol. Immunol. 30, 1587-1592, 1993  
 A>Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic conve  
 A/Reference number: 157824; MUID:94067177; PMID:8247029  
 A/Accession: 157824  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-31; Q, 33-764 <RE2>  
 A/Cross-references: GB:U15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
 C/Comment: 292-Cys has a free sulfhydryl.  
 C/Genetics:  
 A:Gene: GDB:BF  
 A/Cross-references: GDB:119726; OMIM:138470  
 A/Map position: 6p21.3-6p21.3  
 A/Intons: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 69/  
 A>Note: the list of introns may be incomplete  
 A>Note: gene is located in the major histocompatibility complex, class III region  
 C/Complex: complement factor B initially forms an inactive complex with complement factor  
 C/Function:  
 A>Description: Bb is a serine proteinase; C3/C5 convertase cleaves complement C3 alpha ct  
 A1  
 A/Pathway: complement alternate pathway  
 C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; v  
 C/Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydroly  
 P1-25/Domain: signal sequence #status predicted <SIG>  
 P26-764/Product: complement factor B #status experimental <MAT>  
 F26-259/Product: complement factor Ba fragment #status experimental <BAF>  
 F37-98/Domain: complement factor H repeat homology <FH1>  
 F103-158/Domain: complement factor H repeat homology <FH2>  
 F165-218/Domain: complement factor H repeat homology <FH3>  
 F260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
 F268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
 F482-752/Domain: trypsin homology #status atypical <TRY>  
 F37-76; 62-98; 103-145; 131-158; 165-205; 191-218; 478-596; 511-527; 599-615; 656-682; 695-725/Pd  
 F122; 142; 285; 378/Binding site: carbohydrate (asn) (covalent) #status experimental  
 F259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
 F526; 576; 699/Active site: His, Asp, Ser #status experimental

Query Match 11.2%; Score 105.5; DB 1; Length 764;  
 Best Local Similarity 22.6%; Pred. No. 0.34;  
 Matches 48; Conservative 44; Mismatches 75; Indels 45; Gaps 10;

Qy	3	DLVFLIDKSGSV-----LHNMETIYFVQQLAKFISPOLRMFIFSTRGTLMLKTE- 56
Db	270	NIYLVLDGSDSISGASNFTGAKCLVLEKVAISYGRK--RYGLVYVATYPIKIVKVSSEA 327
Qy	57	---DROIQGLLEL-----QKVLPGDQTYWHEGFERSAEQIYYENR-----QGY-RTA 101
Db	328	DSSNADWVTKQLNEIYEDHKLSGNT-----KQLQAVYISMMSWPRDPVEGKMKRTR 381
Qy	102	SVIALLTDG-----ELHEDLFYSREANRNRDGLAIYYCVG--VKDPNETQ 146
Db	382	HVILMTDGLHNMGGDPITVDEIRDLTYIGDKRKPRREDYLDVYVGVGLVNVQVINA 441
Qy	147	LRIIDSKDHNVPVNDGPDALGQIHSILKKS 178

Db 442 LASKDNEQHVFVKD-MENIDVYQWIDES 472

## RESULT 15

821369

collagen alpha 2(VI) chain precursor - mouse

C/Species: Mus musculus (house mouse)

C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004

C/Accession: S21369; S28808; S13745; S18863

R/Author: X.Y.Z.; Bardou, X.Y.Z.; Dani, C.

A/Reference number: S21369

A/Accession: S21369

A/Molecule type: mRNA

A/Residues: 1-1029 &lt;IBR&gt;

A/Cross-references: UNIPROT:Q02788; EMBL:X65582; NID:949808; PIDN:CAA46541.1; PID:949805

R/Author: A.; Bertrand, B.; Bardou, S.; Amel, E.Z.; Grimaldi, P.; Allhaud, G.; Dani, C.

B/Chem. J. 289, 141-147, 1993

A/Title: Cloning of alpha2 chain of type VI collagen and expression during mouse develop

A/Reference number: S28808; MUID:93143659; PMID:8380980

A/Accession: S28808

A/Molecule type: mRNA

A/Residues: 266-1029 &lt;IB2&gt;

A/Cross-references: EMBL:X62332; NID:949906; PIDN:CAA44206.1; PID:949907

R/Author: C.D.; Jimenez, S.A.

Macrix 11, 1-9, 1991

A/Title: Structure of cDNAs encoding the triple-helical domain of murine alpha-2(VI) col

rate oligonucleotides for generation of novel cDNA clones.

A/Reference number: S13745; MUID:91226374; PMID:1709252

A/Accession: S13745

A/Molecule type: mRNA

A/Residues: 266-267, 'S', 269-294, 'L', 296-600 &lt;CON&gt;

A/Cross-references: GB:J06343; NID:9192671; PIDN:AAA37441.1; PID:9192672

A/Note: the sequence from Fig. 3 is inconsistent with that from Fig. 4 in having 306-116

C/superfamily: collagen VI; von Willebrand factor type A repeat homology

F/1-28/Domain: signal sequence #status predicted &lt;SIG&gt;

F/29-1029/Product: collagen alpha 2(VI) chain #status predicted &lt;MAT&gt;

F/29-265/Domain: globular #status predicted &lt;NC2&gt;

F/54-226/Domain: von Willebrand factor type A repeat homology &lt;VWA1&gt;

F/266-600/Domain: collagenous #status predicted &lt;COL&gt;

F/376-378/Region: cell attachment (R-G-D) motif

F/436-438/Region: cell attachment (R-G-D) motif

F/499-501/Region: cell attachment (R-G-D) motif

F/508-510/Region: cell attachment (R-G-D) motif

F/549-551/Region: cell attachment (R-G-D) motif

F/601-1029/Domain: globular #status predicted &lt;NC1&gt;

F/623-799/Domain: von Willebrand factor type A repeat homology &lt;VWA2&gt;

F/841-1012/Domain: von Willebrand factor type A repeat homology &lt;VWA3&gt;

F/150,337,640,907,963/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.24; Score 105.5; DB 1; Length 1029;

Best Local Similarity 23.18; Pred. No. 0.48;

Matches 49; Conservative 41; Mismatches 85; Indels 37; Gaps 8;

QY 3 DLYFIDKSGSV-----LHHMNE-IYFVFOALHKFISPOLRMSE-----IVFSTRG 48  
 DB 56 NVTFVLDTSVAMQSTDLNMQQFPOFISQLQNEFYLDQVALSMRYGGIHFSDQV 115  
 QY 49 TLMKLTEDREQROGLEBLQKVLPGSDTYMHGFEFASBOIYENRQGYRTASVIALT 108  
 DB 116 EVSPSPGSDASFTKSLQGRSFRG--TFTDCLAMNTQIQHVGKGVNFAVVI--T 171  
 QY 109 DGLHEDLPFYSERENRSDLAIVYCVGV-KDFNETOLARIADSODHVPFVN----- 161  
 DB 172 DGHVTSPPCGIKQARAREEGIRLPALAPNRNLNEQGLRIANSFHELYRNRYATMRP 231  
 QY 162 DGFQALQGIHSHIL-----KKSCIE 181  
 DB 232 DSTEIDDDTINRIIKWKMEAYGECKVSCIE 263

Search completed: June 13, 2005, 20:06:38  
 Job time : 13.9858 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 13, 2005, 19:37:47 ; Search time 60.7608 Seconds  
(without alignments)  
1525.429 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_42\_222

Perfect score: 943  
Sequence: 1 GFDLYRILDKSGSVLHWMNE.....DGFQALQGIHSILKSCIE 181

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_03.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	943	100.0	564 1	Q9H6X2 Homo sapien
2	934	99.0	562 1	Q9CZ52 mus musculu
3	542	57.5	489 1	P58335 homo sapien
4	540	57.3	487 2	Q6DFX2
5	375	39.8	641 2	Q8BVM2
6	146	15.5	1332 2	Q9BPQ8
7	141.5	15.0	1161 1	ITAD RAT
8	139	14.7	441 2	Q8T6U5
9	139	14.7	441 2	Q8T6U5
10	136	14.4	453 2	Q8T5C2
11	136	14.4	724 2	Q04588
12	134.5	14.3	3119 1	CA1C MOUSE
13	132.5	14.1	1152 1	ITAM HUMAN
14	129.5	13.7	1182 2	Q8C6K9
15	128	13.6	3124 1	CA1C CHICK
16	127.5	13.5	712 2	Q43981
17	127.5	13.5	765 2	Q9U8U9
18	126.5	13.4	3063 1	CA1C HUMAN
19	125.5	13.3	1163 1	ITAX MOUSE
20	124.5	13.2	1153 1	ITAM MOUSE
21	124.5	13.2	1162 1	ITAD HUMAN
22	122	12.9	517 2	Q438E3
23	121.5	12.9	920 2	Q289E4
24	119	12.6	1626 2	Q8NFW1
25	117.5	12.5	760 2	Q70350
26	117	12.4	1823 2	Q7PRP5
27	116.5	12.4	919 2	Q75R52
28	116	12.3	637 2	Q8IYX1
29	116	12.3	1284 2	Q6PI59
30	116	12.3	1329 1	Q9P218
31	116	12.3	2944 2	Q63870

32	116	12.3	3183 2	Q65ZC2	Q65ZC2 caenorhabdi
33	116	12.3	3767 1	MOA3 CAEL	P41576 caenorhabdi
34	114.5	12.1	760 1	CO2 MOUSE	P21180 mus musculu
35	114.5	12.1	2104 2	Q21281	Q21281 caenorhabdi
36	114.5	12.1	2104 2	Q964R4	Q964R4 caenorhabdi
37	113.5	12.0	790 2	Q6DC06	Q6DC06 xenopus lae
38	113	12.0	599 2	Q8WVQ1	Q8WVQ1 boletia vi
39	113	12.0	1888 1	CA1B CHICK	P32018 gallus gall
40	112	11.9	588 2	Q6ZED8	Q6ZED8 synechocyst
41	112	11.9	652 2	Q95L12	Q95L12 bos taurus
42	111.5	11.8	929 1	CA1C NOTVI	Q91145 notophthalm
43	111	11.8	440 2	Q8CBT2	Q8CBT2 mus musculu
44	111	11.8	848 2	Q8C720	Q8C720 mus musculu
45	111	11.8	3567 2	Q9ES77	Q9ES77 mus musculu

## ALIGNMENTS

RESULT 1  
ATRI\_HUMAN STANDARD: PRT; 564 AA.  
ID ATRI\_HUMAN  
AC Q9H6X2; Q96P02; Q9WVP3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
GN Name=ANTXR1; Synonyms=ATR, TEM8;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20407466; PubMed=10947988; DOI=10.1126/science.289.5482.1197;  
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.B.,  
RA Montgomery E., Lai A., Riggs G.J., Lengauer C., Vogelstein B.,  
RA Kinzler K.W.;  
RT "Genes expressed in human tumor endothelium";  
RL Science 289:1197-1202(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH ANTHRAX TOXIN.  
RX MEDLINE=21557240; PubMed=11700562; DOI=10.1038/n510198;  
RA Bradley K.A., Mogridge J., Mouraz M., Collier R.D., Young J.A.T.;  
RT "Identification of the cellular receptor for anthrax toxin";  
RL Nature 414:225-229(2001).  
RN [3]  
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.  
RP (ISOFORM 3).  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahara K., Murakami K., Yasuda T., Watanabe M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Puruya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida T.,  
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togawa S., Komai F., Haru R., Takeuchi K., Arita M.,  
RA Imose N., Mueshino K., Yuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nishigawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujitara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroto M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hata H., Watanabe Y., Komatsu T.,  
RA Mizushima-Sugano J., Setoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RT Nat. Genet. 36:40-45(2004).  
RN [4]  
RN SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Kidney;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,  
RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Tsibiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mulhally S.J.,  
RA Bosak S.A., McEwan P.V., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smillie D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RL [5]  
RL INTERACTION WITH ANTHRAX TOXIN.  
RN [5]  
RN TISSUE=Placenta;  
RC MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
RX Scobie H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.,  
RA "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
RA receptor.";  
RA Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
RN [6]  
RN PUBMED=14759258; DOI=10.1186/gb-2004-5-2-r8;  
RX Hillman R.T., Green R.E., Brenner S.E.,  
RA "An unappreciated role for RNA surveillance.";  
RL Genome Biol. 5:RESEARCH008.1-RESEARCH008.16(2004).  
CC -1- FUNCTION: Cellular role is not yet known.  
CC -1- SUBUNIT: Binds to the protective antigen (PA) of *Bacillus*  
CC anthracis. Binding does not occur in the presence of calcium.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Comment=Experimental confirmation may be lacking for some  
CC isoforms;  
CC Name=1;  
CC IsoId=Q9H6X2-1; Sequence=Displayed;  
CC Note=May be produced at very low levels due to a premature stop  
CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
CC Name=2;  
CC IsoId=Q9H6X2-2; Sequence=VSP\_000444, VSP\_000445;  
CC Name=3;  
CC IsoId=Q9H6X2-3; Sequence=VSP\_000446, VSP\_000447;  
CC Name=4;  
CC IsoId=Q9H6X2-4; Sequence=VSP\_000448, VSP\_000449;  
CC -1- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells  
CC but not in normal endothelial cells.  
CC -1- DOMAIN: Binding to PA seems to be effected through the VMA domain.  
CC -1- SIMILARITY: Belongs to the ATR family.  
CC -1- SIMILARITY: Contains 1 VMPA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF279145; AAK52094.1; -;  
DR EMBL; AF421380; AAL26496.1; -;  
DR EMBL; AK025429; BAB15128.1; ALT INIT.  
DR EMBL; AK001463; BAA91707.1; ALT\_FRAME.  
DR EMBL; BC012074; AAI12074.1; -;  
DR Genew; HGNC:21014; ANTXRL.  
DR H-InvDB; HIX0002125; -;  
DR MIM; 606410; -;  
DR InterPro; IPR008400; Anth\_Ig.  
DR InterPro; IPR008399; Ant\_C.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF05587; Anth\_Ig; 1.  
DR Pfam; PF05586; Ant\_C; 1.  
DR Pfam; PR00982; VMA; 1.  
DR SMART; SM00327; VMA; 1.  
DR PROSITE; PS50234; VMPA; 1.  
KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
FT CHAIN 1 32  
FT SIGNAL 1 32  
FT DOMAIN 33 321  
FT TRANSMEM 322 342  
FT DOMAIN 343 564  
FT DOMAIN 44 215  
FT DOMAIN 360 368  
FT DOMAIN 506 564  
FT CARBOHYD 166 166  
FT CARBOHYD 184 184  
FT CARBOHYD 262 262  
FT VARSPPLIC 365 368  
FT VARSPPLIC 369 564  
FT VARSPPLIC 268 297  
FT VARSPPLIC 298 564  
FT VARSPPLIC 319 333  
FT VARSPPLIC 334 564  
SQ SEQUENCE 564 AA; 62789 MW; B118A00D5DF2233 CRC64;  
Query Match 100.0%; Score 943; DB 1; Length 564;  
Best Local Similarity 100.0%; Pred. No. 9, 2e-73;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GPDYFILDKSGSVLHNNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTMLKLTBREQ 60  
DB 42 GPDYFILDKSGSVLHNNMEIYFVEQLAHKFI SPOLRMSFIYSTRGTTMLKLTBREQ 101  
QY 61 IRQGELKQVLPFGSDTYNHEGFERASEQIYYENRGYRTASYIALTDGELHEDLPFYS 120  
DB 102 IRQGELKQVLPFGSDTYNHEGFERASEQIYYENRGYRTASYIALTDGELHEDLPFYS 161  
QY 121 EREANRSRLGALIVYCVGVKDNFETOLARIADSKDHVPVNDGFQALOGIHSILKKSCT 180  
DB 162 EREANRSRLGALIVYCVGVKDNFETOLARIADSKDHVPVNDGFQALOGIHSILKKSCT 221  
QY 181 E 181  
DB 222 E 222  
RESULT 2  
ATRL\_MOUSE STANDARD; PRT; 562 AA.



RT membrane matrix assembly, cell cycle progression, cellular  
RT differentiation and G-protein signaling.";  
RL U. Cell Sci. 114:2755-2773 (2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH ANTHRAX TOXIN.  
RC TISSUE=Placenta; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
R MEDLINE=22608610;  
RA Scoble H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.;  
RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
receptor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174 (2003).  
RN [3]  
RP SEQUENCE OF 78-489 FROM N.A. (ISOFORM 3), AND SEQUENCE FROM N.A.  
R (ISOFORM 4).  
RC TISSUE=Synovial cell;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Odayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto U.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y.,  
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hottu T.,  
RA Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
RA Iose N., Masehino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunaga H., Ichihara T., Shiohara T., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Tereshima T., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohnori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Sena T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama H., Hata H., Watanabe M., Komatsu N.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamaeita R.,  
RA Nakai K., Ieda T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
cDNAs.";  
RL Nat. Genet. 36:40-45 (2004).  
RN [4]  
RP Nat. Genet. 36:40-45 (2004).  
RT [4]  
CC -1- FUNCTION: Cellular role is not yet known.  
CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
CC anthracis in a divalent cation-dependent manner, with the  
CC following preference: calcium > manganese > magnesium > zinc.  
CC Seems to bind to collagen type IV and laminin.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1 and 2).  
CC Secreted (isoform 3). Isoform 1 is expressed at the cell surface  
CC while isoform 2 is predominantly expressed within the endoplasmic  
CC reticulum and not at the plasma membrane.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named Isoforms=4;  
CC Name=1;  
CC IsoId=P58335-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=P58335-2; Sequence=VSP\_008343;  
CC Name=3;  
CC IsoId=P58335-3; Sequence=VSP\_008344, VSP\_008345;  
CC Note=No experimental confirmation available;  
CC Name=4;  
CC IsoId=P58335-4; Sequence=VSP\_008346;  
CC Note=No experimental confirmation available;  
CC -1- TISSUE SPECIFICITY: Expressed in colon, heart, kidney, lung,  
CC liver, peripheral blood leukocytes, placenta, skeletal muscle,  
CC small intestine and spleen.  
CC -1- DOMAIN: Binding to PA seems to be effected through the VMA domain.  
CC -1- SIMILARITY: Belongs to the ATR family.

CC -1- SIMILARITY: Contains 1 VMA domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
CC -----  
DR EMBL; AY040326; AAK7222.1; -;  
DR EMBL; AY233452; AAK04016.1; -;  
DR EMBL; AK055636; BAB70976.1; ALT\_INIT.  
DR EMBL; AK091721; BAC03731.1; -;  
DR Genew; HGNC:21732; ANTXR2.  
DR MIM; 608041; -;  
DR Interpro; IPR008399; Ant C.  
DR Interpro; IPR002035; VMA\_A.  
DR Pfam; PF05586; Ant\_C\_1.  
DR Pfam; PF00092; VMA\_1.  
DR PROSITE; PS50234; VMAF\_1.  
KW Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 33  
FT CHAIN 34 489  
FT DOMAIN 34 318  
FT TRANSMEM 319 341  
FT DOMAIN 342 489  
FT DOMAIN 44 213  
FT CARBOHYD 250 250  
FT CARBOHYD 260 260  
FT VARSPPLIC 213 315  
FT  
FT VARSPPLIC 290 322  
FT  
FT VARSPPLIC 323 489  
FT  
FT VARSPPLIC 477 489  
FT  
FT CONFLICT 357 357 P -> A (in Ref. 3; BAC03731).  
FT SQ SEQUENCE 489 AA; 53692 MW; B9F679DB7586E2B7 CRC64;  
Query Match 57.5%; Score 542; DB 1; Length 489;  
Best Local Similarity 59.4%; Pred. No. 2,7e-38;  
Matches 107; Conservative 35; Mismatches 36; Indels 2; Gaps 1;  
QY 2 FDIYFLDSSGSLVHNMNRYVFEQLAKFIPOLRMSTIVSTGKTLMLKTEDEQI 61  
DB 43 FDIYFLDSSGSLVHNMNRYVFEQLAKFIPOLRMSTIVSTGKTLMLKTEDEQI 102  
QY 62 RGLLEELQKLVPGGDIYVMEHGFASQIYENRGKRTASVIALTDGELHEDLFFYSE 121  
DB 103 SKLELDKRVSPGERTYIHGLKLANEQI-QVAGLKTSSIIITLTDGLDGLVPSYAE 160  
QY 122 REANRSRDIGAIYVYGVNDPMTQARIADSKDHVFPVNDGFQALQGIHSLIKSCIE 181  
DB 161 KEAKISRSIGASVYCVGLDFEQALRIADSKDHVFPVNDGFQALQGIHSLIKSCIE 220  
RESULT 4  
Q6DFX2  
ID Q6DFX2 PRELIMINARY; PRT; 487 AA.  
AC Q6DFX2;  
DT 25-OCT-2004 (T-EMBLrel. 28, Created)  
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)  
DE Anthrax toxin receptor 2.  
GN Name=Antxr2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;  
 OX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Krausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Krausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Datchenko L., Marusik A., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Tothiyuki S., Carninci P., Prange C.T.,  
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bobak S.A., McEwan P.U., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pheyl J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzywiński M.I., Skalek U., Smallie D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RA Krausberg R.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC076595; AAF76595.1; -  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0004872; F: receptor activity; IEA.  
 DR InterPro: IPR008400; Anth Ig.  
 DR InterPro: IPR008399; Ant C.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF05587; Anth\_Ig; 1.  
 DR Pfam: PF05586; Ant\_C; 1.  
 DR Pfam: PF00092; VMA; 1.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PSS0234; VMA; 1.  
 KM Receptor.  
 SQ SEQUENCE 487 AA; 53184 MW; 61A40D60BC8DE69 CRC64;  
 Query Match 57.3%; Score 540; DB 2; Length 487;  
 Best local similarity 59.4%; Pred. No. 4e-38;  
 Matches 107; Conservative 32; Mismatches 39; Indels 2; Gaps 1;  
 QY 2 FDLYFLDSDSGSVLHNNELIYFVEQLAHKFTSPQLMSFIVSTGCTIMKTEDREOI 61  
 DB 43 FDLYFLDSDSGSVANMWIEIYFNHQLTERFVSEKMLSTFVSSQNTIILPLTGDRYKI 102  
 QY 62 RQLELELQKLVPGGDTYTMHGGFERASQIYENKQRTKTSVITALTGELHEDLPFYSE 121  
 DB 103 GKLELELQKLVPGGDTYTMHGGFERASQIYENKQRTKTSVITALTGELHEDLPFYSE 160  
 QY 122 REANRSRDLCATYTCVGVKDFENETQLARIADSKDHPVAVDQFALOGIHSITKSCIE 181  
 DB 161 NEAKRSRLGASVYCGVLDPEQQLERLADSKDQVFPVKGFOALGKIIINSLAOSCTE 220  
 RESULT 5  
 Q8BVM2 PRELIMINARY; PRT; 641 AA.  
 AC Q8BVM2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched  
 DE library, clone:493430J11 product:hypothetical Prolin-rich region/von  
 DE Willebrand factor type A domain containing protein, full insert  
 DE sequence.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RA The FANTOM Consortium;  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kikunishi T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Testis;  
 RA Aichi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hamagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Kato F., Imotani K., Ishii Y., Itoh M., Kigawa I., Kasukawa T.,  
 RA Kich H., Kawai J., Kojima Y., Kondo S., Kono H., Koyama M., Koya S.,  
 RA Kishihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ono M., Ohnato N., Okazaki Y.,  
 RA Saito K., Saitoh H., Sakai C., Sakai K., Sakatsume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami T.,  
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Taya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK077206; BAC36683.1; -  
 DR GO: GO:0016021; C: integral to membrane; IEA.  
 DR GO: GO:0004872; F: receptor activity; IEA.  
 DR InterPro: IPR008400; Anth Ig.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF05587; Anth\_Ig; 1.  
 DR Pfam: PF00092; VMA; 1.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PSS0234; VMA; 1.

KM Hypothetical protein.  
SQ SEQUENCE 641 AA; 70415 MW; 1995300730BC8553 CRC64;

Query Match 39.8%; Score 375; DB 2; Length 641;  
Best Local Similarity 43.9%; Pred. No. 9e-24;  
Matches 79; Conservative 31; Mismatches 70; Indels 0; Gaps 0;

QY 2 FDLVFLDKSGSVLHNNELIYFVEQLAKRFPISQLMSFVSTRTGTLMLKLTEDREQ 61  
DB FDLVFLDKSGSVADNMIHLYSFAELGVKKFTNPMLISITVSTAEVLLPLTSDKEI 134

QY 62 RQGLEELQKVLPGDVTWHEGFERASEQIYENNGYRTASVITATDGEIHEDLFPYSE 121  
DB 135 NKSILVTKSVIPQGLTMMQKLRKANEOIKSTIGRIANSVITATDGLLLKPYIDTM 194

QY 122 REANRSDLAIVYCVGKDFNETOLARIADSKDHPVNDGFOALQGIHISILKSCIR 181  
DB 195 EEAKKARMGALVYTVGVFMYSKQVLNIGDPDRCGVDEGFALEGVVDPLTSKSCTE 254

RESULT 6  
Q9BP08 PRELIMINARY; PRT; 1332 AA.

AC Q9BP08;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE Integrin alpha H1 precursor.  
GN Name=H1RGAL;  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Scolidobranchia; Pyuridae; Halocynthia.  
OX NCBI\_TaxID=7723;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Hemocyte;  
RA MEDLINE=21103187; Pubmed=11160215;  
RX Miyazawa S., Azumi K., Nonaka M.;  
RT "Cloning and characterization of integrin alpha subunits from the  
RT solitary ascidian, Halocynthia roretzi.";  
RL J. Immunol. 166:1710-1715 (2001).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
DR EMBL; AB048261; BAB21479.1; -.  
DR HSSP; P11215; 1BHQ.  
DR GO; GO:0016021; C:Integral to membrane; IEA.  
DR GO; GO:0008305; C:Integrin complex; IEA.  
DR GO; GO:0005515; F:Protein binding; IEA.  
DR GO; GO:0007160; P:Cell-matrix adhesion; IEA.  
DR GO; GO:0007229; P:Integrin-mediated signaling pathway; IEA.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00092; VMA; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VMA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KM Cell adhesion; Integrin; Signal; Transmembrane.  
FT SIGNAL 1 30 Potential.  
FT CHAIN 31 1332 Integrin alpha H1.  
SQ SEQUENCE 1332 AA; 145851 MW; 0D9108D2B05CFEAB CRC64;

Query Match 15.5%; Score 146; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 0.0011;  
Matches 52; Conservative 34; Mismatches 60; Indels 66; Gaps 7;

QY 1 GFLPLFLDKSGSVLHNNELIYFVEQLAKRFPISQLMSFVSTRTGTLMLKLTEDREQ 60  
DB 204 GAVVFLFLDKSGSVGKRFDKVKKVKNIT-----AKLDIGKEI 241

QY 61 IROGLEELQKVLPGDVTWHEGFERASEQIYENNGYRT----- 100

DB 242 VRVGVQYSHYVSGSKINKQYITTEISIGFELDNFENAVDRI---QIQGYTTYTGRA 298

QY 101 -----ASVITATDGEIHEDLFPYSEREANRSDLAIVYCVGKDFNET 145  
DB 299 LQKVRIDPDPAVIGNKQVLLLTLDGQAKDKILP--NANRLNKGATPAVGVGEYDIS 356

QY 146 QLARIA---DSKDHFPVNDGFOALQGIHISI 174  
DB 357 ELKLIASGTDSTDRVFTVD-FGEIDLSIVKSL 387

RESULT 7  
ITRD\_RAT  
ID ITRD\_RAT STANDARD; PRT; 1161 AA.

AC Q9QYE7;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Integrin alpha-D precursor.  
GN Name=Itgad;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RA O'Brien M.M., VanderVlieten M., Kilgannon P.D., Dietach G.,  
RA Gallatin W.M.;  
RT "Cloning of rat alpha D, a novel beta 2 integrin."  
RL Submitted (Aug-1997) to the EMBL/GenBank/DBD databases.  
CC -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICM3 and VCAM1. May play a role in the atherosclerotic process such as clearing lipoproteins from plaques and in phagocytosis of blood-borne pathogens, particulate matter, and senescent erythrocytes from the blood (By similarity).  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D associates with beta-2 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).  
CC -1- DOMAIN: The integrin I-domain (insert) is a VMA domain. Integrins with I-domains do not undergo protease cleavage.  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -1- SIMILARITY: Contains 1 VMA domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to ilcense@isb-sib.ch) or send an email to ilcense@isb-sib.ch).  
DR EMBL; AF021334; AAP21241.1; -.  
DR HSSP; P11215; 1BHQ.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01839; FG-GAP; 3.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR Pfam; PF00092; VMA; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VMA; 1.  
DR PROSITE; PS50242; INTEGRIN\_ALPHA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KM Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
FT Repeat; Signal; Transmembrane.  
FT SIGNAL 1 19 Potential.  
FT CHAIN 20 1161 Integrin alpha-D.  
FT DOMAIN 20 1100 Extracellular (Potential).  
FT TRANSMEM 1101 1121 Potential.  
FT DOMAIN 1122 1161 Cytoplasmic (Potential).

```
FT REPEAT 34 87 FG-GAP 1.
FT REPEAT 88 7 FG-GAP 2.
FT DOMAIN 152 334 VFMA.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.
FT CA_BIND 467 475 Potential.
FT CA_BIND 531 539 Potential.
FT CA_BIND 594 602 Potential.
FT SITE 1126 1130 GFPKR motif.
FT DISULFID 69 76 By similarity.
FT DISULFID 108 126 By similarity.
FT DISULFID 656 711 By similarity.
FT DISULFID 769 775 By similarity.
FT DISULFID 845 860 By similarity.
FT DISULFID 993 1017 By similarity.
FT DISULFID 1022 1027 By similarity.
FT CARBOHYD 61 61 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 89 89 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 245 245 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 393 393 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 696 696 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 734 734 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 784 784 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 907 907 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 936 936 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 1045 1045 N-linked (GlcNAc. . .) (Potential).
SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;
```

Query Match 15.0%; Score 141.5; DB 1; Length 1161;  
Best Local Similarity 27.7%; Pred. No. 0.0023;  
Matches 53; Conservative 38; Mismatches 75; Indels 25; Gaps 10;

```
QY 3 DLVFIIDKSSGV-LHNNNEIYFVEQLAHKF-ISPOLMSFVFSRGTTLMLKT----- 56
DB 152 DIAFLIDGSGINORDPAQKMDFKALMGEPASTSLFSLMOYSNLIKHTFTFEFNIL 211
QY 57 DRAQIRQGLELOKVLPGGDTYMHGEPERASEQIYYENRQGYRTA-SVITALTDELHED 115
DB 212 DPGSLVDPIYQLQ-----GITTYATGIRTYMEELFHSKNSRKSAAKILVIITDGKTRD 266
QY 116 LFPEYSR--REANRSRDGAIVYCVGV-D-FNE-TOLARI-----ADSKDHVPVNDGFOA 166
DB 267 PLEYSDVTPADADA---GIRYALGVGDARFQEPALMLWTIGSAPRQDHVFKYGN-FAA 322
QY 167 LQGIHSLIKK 177
DB 323 LRSLOROLQEK 333

RESULT 8
Q8T6U5 PRELIMINARY; PRT; 441 AA.
ID Q8T6U5;
AC Q8T6U5;
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Proximal thread matrix protein 1 variant a.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_Taxid=6550;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=22313036; Pubmed=12425661; DOI=10.1021/bm0255903;
RX Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric exteroorganismic
RT bysael fibers";
RL Biomacromolecules 3:1240-1248(2002).
DR EMBL; AF14454; AAL83537.1; -
DR GO; GO:0005198; F:structural molecule activity; IEA.
```

```
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VFMA; 2.
KW Matrix protein.
SQ SEQUENCE 441 AA; 47543 MW; 881DBBD36B8912B CRC64;
```

Query Match 14.7%; Score 139; DB 2; Length 441;  
Best Local Similarity 25.1%; Pred. No. 0.0012;  
Matches 49; Conservative 46; Mismatches 74; Indels 26; Gaps 12;

```
QY 3 DLVFIIDKSSGV-LHNNNEIYFVEQLAHKF-ISPOLMSFVFSRGTTLMLKT 55
DB 241 DIAVFDPASSINANNNNYGLMKDVFMDIYDRFNKTPGPGTQPAVVTFAADRATKQFGLK 300
QY 56 E--DREQIRQGLELOKVLPG--GDTYMHGEPERASEQIYYENRQ---YRTASVIALT 108
DB 301 DYSSKAIRKAIID---KYTSIIIGQTALIGDLENARLEV-FPNRNGGGRREVQKVVILLT 356
QY 109 DGBL--HEDLFYSEERANRSRDGAIVYCVGV-KDENEQTOLARIADSKDHVPVNDGFOQ 165
DB 357 DGQNNGHKS----PEHSSILRKRGVYIVAIQVGTGFLKSELINIASSEYVF-TTSSFN 411
QY 166 ALQGIHSLIKKSCI 180
DB 412 KLSKIMENVVYLAQM 426
```

## RESULT 9

Q8T5C3 PRELIMINARY; PRT; 444 AA.

```
ID Q8T5C3;
AC Q8T5C3;
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Proximal thread matrix protein 1b.
OS Mytilus edulis (Blue mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_Taxid=6550;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=22313036; Pubmed=12425661; DOI=10.1021/bm0255903;
RX Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric exteroorganismic
RT bysael fibers";
RL Biomacromolecules 3:1240-1248(2002).
DR EMBL; AY053390; AAL17973.1; -
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VFMA; 2.
KW Matrix protein.
SQ SEQUENCE 444 AA; 47615 MW; D2C605347450C931 CRC64;
```

Query Match 14.7%; Score 139; DB 2; Length 444;  
Best Local Similarity 25.1%; Pred. No. 0.0012;  
Matches 49; Conservative 46; Mismatches 74; Indels 26; Gaps 12;

```
QY 3 DLVFIIDKSSGV-LHNNNEIYFVEQLAHKF-ISPOLMSFVFSRGTTLMLKT 55
DB 244 DIAVFDPASSINANNNNYGLMKDVFMDIYDRFNKTPGPGTQPAVVTFAADRATKQFGLK 303
QY 56 E--DREQIRQGLELOKVLPG--GDTYMHGEPERASEQIYYENRQ---YRTASVIALT 108
DB 304 DYSSKAIRKAIID---KYTSIIIGQTALIGDLENARLEV-FPNRNGGGRREVQKVVILLT 359
QY 109 DGBL--HEDLFYSEERANRSRDGAIVYCVGV-KDENEQTOLARIADSKDHVPVNDGFOQ 165
DB 360 DGQNNGHKS----PEHSSILRKRGVYIVAIQVGTGFLKSELINIASSEYVF-TTSSFN 414
```

```
Qy      166 ALOGIHSLKSC1 180
      | : : : : :
Db      415 KLSKIMENVKLCM 429

RESULT 10
Q875C2
ID Q875C2 PRELIMINARY; PRT; 453 AA.
AC Q875C2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Proximal thread matrix protein 1.
OS Mytilus galloprovincialis (Mediterranean mussel).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_TaxID=29158;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;
RA Sun C., Lucas J.M., Waite J.H.;
RT "Collagen-binding matrix proteins from elastomeric extraorganic
RT lysal fibers.";
RL Biomacromolecules 3:1240-1248(2002).
DR HMBL; AY053391; AAL17974.1; -.
DR HSBP; P20701; IMCN.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00092; VMA; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 2.
DR PROSITE; PS50234; VWF_A; 2.
DR Matrix protein.
KW SEQUENCE 453 AA; 48784 MW; D60497F5C0C51BED CRC64;
SQ

Query Match 14.4%; Score 136; DB 2; Length 453;
Best Local Similarity 25.1%; Pred. No. 0.0023;
Matches 49; Conservative 44; Mismatches 76; Indels 26; Gaps 12;

Qy      3 DLVEIILDKSGSV---LHMNEIYFVEQLAHKF--ISPQ-LRMSFIVFSTRTGLTKLT 55
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     253 DIAFVFPASSINANNNNVQLMKNFMDIVDRNPKTGPGTGPVAVTFADRAIKRGK 312
Qy      56 E--DREOIRQGLELOVLP--GDTYMHGFEFASQIYENRQ--YRTASVITALT 108
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     313 DYSSKADIKKAID--KVSPSIIGQTALGDLGNARLEV-FPNNGGGRREVQKVILLT 368
Qy      109 DGEI--HEDLFYSEERANSRDGAIYCVGV-KDENTQGLARIADSKDHVPYNDG 165
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     369 DGGNGHKS---PEHSSSLRKGVVIVAGVGTGFLKSELINIASSEEVF-TTISPD 423
Qy      166 ALOGIHSLKSC1 180
      | : : : : :
Db     424 KLSKIMENVKLCM 438

RESULT 11
Q04588
ID Q04588 PRELIMINARY; PRT; 724 AA.
AC Q04588;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Major antigen homologous sequence (emp100).
OS Eimeria maxima.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5804;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=3119203; PubMed=8426611; DOI=10.1016/0166-6851(93)90255-V;
RA Paganontes L.E., Hug D., Huenbein M., Weber G.;
```

```
RT      "Sequence of a major Eimeria maxima antigen homologous to the Eimeria
RT tenella microneme protein Exp100."
RL Mol. Biochem. Parasitol. 57:171-174(1993).
DR EMBL; M99058; AAA29076.1; -.
DR PIR; A48569; A48569.
DR HSBP; P07996; 1LSL.
DR GO; GO:0008083; F:growth factor activity; IEA.
DR InterPro; IPR000762; PTN MK.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00090; TSP_1; 5.
DR Pfam; PF00092; VMA; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS50092; TSP1; 5.
DR PROSITE; PS50234; VWF_A; 1.
SQ SEQUENCE 724 AA; 75808 MW; AC2A0E7A346A7E9E CRC64;

Query Match 14.4%; Score 136; DB 2; Length 724;
Best Local Similarity 27.0%; Pred. No. 0.004;
Matches 53; Conservative 36; Mismatches 79; Indels 28; Gaps 12;

Qy      3 DLVEIILDKSGSV-LHMNEIYFVEQLAHKF-ISP-QLRMSFIVFSTRTGLTKLTEDRE 59
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     47 DWLVVVDGSGISGYSYKGRSIFSNFAGTMPLSPDVGVLVFTGSATVRDLSRSRA 106
Qy      60 QIRQGLEIOLKVP--GGDTYMHGFEFASQIYENRQGR--TASVITALTDEIHED 115
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     107 QNADLLAAAKKLPVAASTYTHGLAKA-EELIFSGKGRDNPAMILVMTDGA----- 161
Qy      116 LFYSERE-----AKSRDLGAIYCVGV-KQNTQLARIA--DSKHV-PP--VNDG 163
      | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db     162 ---SSRRSGTSAAKLRNRGVIVAGVGVNSAECSRIAGCDTVECPRYQSN 217
Qy      164 FQALOGIHSILKSC 179
      | : : : : :
Db     218 WGVSSQINGITIKAC 233

RESULT 12
CALC MOUSE
ID CALC MOUSE STANDARD; PRT; 3119 AA.
AC Q60847; P70322;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(XII) chain precursor.
GN Name=Coll2a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND
RP XIIB-1).
RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
RX MEDLINE=96170761; PubMed=8601036;
RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
RT "Primary structure of the long and short splice variants of mouse
RT collagen XII and their tissue-specific expression during embryonic
RT development.";
RL Dev. Dyn. 204:432-445(1995).
RN [2]
RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2
RP AND XIIB-2).
RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
RX MEDLINE=99348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;
RA Kania A.W., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
RA Olsen B.R., Nishimura I.;
RT "Structural variation of type XII collagen at its carboxyl-terminal
RT NCI domain generated by tissue-specific alternative splicing.";
RL J. Biol. Chem. 274:22053-22059(1999).
```





Matches 56; Conservative 37; Mismatches 74; Indels 23; Gaps 11;

QY 3 DLVFLDKSGSV-LHNNIEIYFVEQLAKF-ISP-QLRMSFVPSRGTTLMLKLTEDRE 59  
 DB 444 DIVFLVGSYSIGTANFKVAFLEVLAKSPFISPNRVQSLVSGSDPHTEFTLKFNRR 503  
 QY 60 QIRGCELEQLKV-----PGSDTYMHGFEFASQIYENRQGRRT--ASTITLTTGEL 112  
 DB 504 ----VEDITIKAITFPFPGGNTGTAKMTYRREKIVPNK-GSRSNVPMKMLITDGE- 556  
 QY 113 HEDLFFSEBEANRSDLGAIIVYCVGVDFNETQLARIAD--SKDHYFPVNDGFOALQGI 170  
 DB 557 SSDAF---RDPALTLRNSDVEIFAVGVKDVANRSLKLAISAPATHYFTVED-FDAQORI 612  
 QY 171 IHSILKKSCI 180  
 DB 613 SFVLTSICL 622

RESULT 13  
 ITAM\_HUMAN STANDARD; PRT; 1152 AA.  
 ID 01-JUN-1989 (Rel. 11, Created)  
 AC P11215;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1) (Neutrophil adherence receptor).  
 DE Name:ITGAM; Synonyms:CD11B, CR3A;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=88315033; PubMed=2457584;  
 RX Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
 RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the Integrins, von Willebrand factor and factor B.";  
 RT J. Biol. Chem. 263:12403-12411 (1988).  
 RN [2]  
 RP MEDLINE=88190151; PubMed=2833753;  
 RX Arnaout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;  
 RT "Molecular cloning of the alpha subunit of human and guinea pig leukocyte adhesion glycoprotein MO1: chromosomal localization and homology to the alpha subunits of Integrins".  
 RT Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780 (1988).  
 RN [3]  
 RP MEDLINE=88257215; PubMed=2454931; DOI=10.1083/jcb.106.6.2153;  
 RX Arnaout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
 RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor MO1 (complement receptor type 3).";  
 RT J. Cell Biol. 106:2153-2158 (1988).  
 RN [4]  
 RP MEDLINE=93123748; PubMed=8419480;  
 RX Fleming J.C., Pahl H.L., Gonzales D.A., Smith T.F., Tenen D.G.;  
 RT "Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-integrin gene family demonstrate remarkable conservation of genomic organization and suggest early diversification during evolution.";  
 RT J. Immunol. 150:480-490 (1993).  
 RN [5]  
 RP MEDLINE=99098893; PubMed=256162;  
 RX Hickey D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L., Roth G.J.;  
 RT "cDNA sequence for the alpha M subunit of the human neutrophil adherence receptor indicates homology to integrin alpha subunits.";

RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261 (1989).  
 RN [6]  
 RP MEDLINE=92073318; PubMed=1683702;  
 RX Shelley C.S., Arnaout M.A.;  
 RT "The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529 (1991).  
 RN [7]  
 RP MEDLINE=8076671; PubMed=3539202; DOI=10.1016/0167-4838(86)90037-3;  
 RX Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaout M.A.;  
 RT "N-terminal sequence of human leukocyte glycoprotein MO1: conservation across species and homology to platelet IIb/IIIa.";  
 RT Biochim. Biophys. Acta 874:368-371 (1986).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-331.  
 RX MEDLINE=95171458; PubMed=7867070; DOI=10.1016/0092-8674(95)90517-0;  
 RA Lee J.O., Rieu P., Arnaout M.A., Liddington R.;  
 RT "Crystal structure of the A domain from the alpha subunit of integrin CD11b/CD18.";  
 RT Cell 80:631-638 (1995).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
 RX MEDLINE=96363671; PubMed=8747460; DOI=10.1016/S0969-2126(01)00271-4;  
 RA Lee J.O., Bankston L.A., Arnaout M.A., Liddington R.;  
 RT "Two conformations of the integrin A-domain (I-domain): a pathway for activation?";  
 RT Structure 3:1333-1340 (1995).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
 RX MEDLINE=9832595; PubMed=9687375; DOI=10.1016/S0969-2126(98)00093-8;  
 RA Baldwin E.T., Saver R.W., Bryant G.L. Jr., Curry K.A., Fairbank M.B., Finzel B.C., Garlick R.L., Heinrichson R.L., Horton N.C., Kelley L.L., Milder A.M., Moon J.B., Mott J.E., Mutchler V.T., Tomich C.S., Watcempaugh K.D., Wiley V.H.;  
 RT "Cation binding to the integrin CD11b I domain and activation model assessment.";  
 RT Structure 6:923-935 (1998).  
 RN [12]  
 RP 3D-STRUCTURE MODELING OF 17-616.  
 RX MEDLINE=98226734; PubMed=9560195; DOI=10.1073/pnas.95.9.4870;  
 RA O'Kvieg C., Springer T.A.;  
 RT "Experimental support for a beta-propeller domain in integrin alpha-subunits and a calcium binding site on its lower surface.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875 (1998).  
 CC -1- FUNCTION: Integrin alpha-M/beta-2 is implicated in various adhesive interactions of monocytes, macrophages and granulocytes as well as in mediating the uptake of complement-coated particles. It is identical with CR-3, the receptor for the IC3b fragment of the third complement component. It probably recognizes the R-G-D peptide in C3b. Integrin alpha-M/beta-2 is also a receptor for fibrinogen, factor X and ICAM1. It recognizes p1 and p2 peptides of fibrinogen gamma chain.  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-M associates with beta-2.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and granulocytes.  
 CC -1- DOMAIN: The integrin I-domain (inset) is a VWFA domain. Integrins with I-domains do not undergo protease cleavage.  
 CC -1- SIMILARITY: Belongs to the integrin alpha chain family.  
 CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC -1- DATABASE: NAME=PROM; NOTE=CD guide CD11b entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11b.htm".

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; J03925; AAA59544.1; -;  
 DR EMBL; M18044; AAA59491.1; -;  
 DR EMBL; J04145; AAA59903.1; -;  
 DR EMBL; S52227; AAB24821.1; -;  
 DR EMBL; S52152; AAB24821.1; JOINED.  
 DR EMBL; S52153; AAB24821.1; JOINED.  
 DR EMBL; S52154; AAB24821.1; JOINED.  
 DR EMBL; S52155; AAB24821.1; JOINED.  
 DR EMBL; S52157; AAB24821.1; JOINED.  
 DR EMBL; S52159; AAB24821.1; JOINED.  
 DR EMBL; S52163; AAB24821.1; JOINED.  
 DR EMBL; S52164; AAB24821.1; JOINED.  
 DR EMBL; S52165; AAB24821.1; JOINED.  
 DR EMBL; S52167; AAB24821.1; JOINED.  
 DR EMBL; S52169; AAB24821.1; JOINED.  
 DR EMBL; S52170; AAB24821.1; JOINED.  
 DR EMBL; S52173; AAB24821.1; JOINED.  
 DR EMBL; S52174; AAB24821.1; JOINED.  
 DR EMBL; S52180; AAB24821.1; JOINED.  
 DR EMBL; S52181; AAB24821.1; JOINED.  
 DR EMBL; S52184; AAB24821.1; JOINED.  
 DR EMBL; S52189; AAB24821.1; JOINED.  
 DR EMBL; S52191; AAB24821.1; JOINED.  
 DR EMBL; S52192; AAB24821.1; JOINED.  
 DR EMBL; S52203; AAB24821.1; JOINED.  
 DR EMBL; S52212; AAB24821.1; JOINED.  
 DR EMBL; S52213; AAB24821.1; JOINED.  
 DR EMBL; S52216; AAB24821.1; JOINED.  
 DR EMBL; S52219; AAB24821.1; JOINED.  
 DR EMBL; S52220; AAB24821.1; JOINED.  
 DR EMBL; S52221; AAB24821.1; JOINED.  
 DR EMBL; S52222; AAB24821.1; JOINED.  
 DR EMBL; S52226; AAB24821.1; JOINED.  
 DR EMBL; M76724; AAA58410.1; -;  
 DR EMBL; M84477; AAA58410.1; -;  
 DR PIR; A31108; RWHU1B.  
 DR PDB; 1A8x; Model; @-17-1152.  
 DR PDB; 1BHO; X-ray; 1/2-  
 DR PDB; 1BHQ; X-ray; 1/2-  
 DR PDB; 1IDN; X-ray; 1/2-  
 DR PDB; 1IDO; X-ray; @-140-331.  
 DR PDB; 1JLM; X-ray; @-143-334.  
 DR PDB; 1MLU; X-ray; A-137-331.  
 DR PDB; 1MF7; X-ray; A-144-337.  
 DR PDB; 1N92; X-ray; A-140-335.  
 DR PDB; 1NA5; X-ray; A-144-345.  
 DR Genew; HGNC:6149; ITGAM.  
 DR MIM; 120980; -;  
 DR CO; GO:0008305; C:Integrin complex; TAS.  
 DR GO; GO:0007155; P:cell adhesion; TAS.  
 DR Interpro; IPR000413; Integrin\_alpha.  
 DR Interpro; IPR02035; VWF\_A.  
 DR Pfam; PF01839; FG-GAP; 3-  
 DR Pfam; PF00357; Integrin\_alpha; 1.  
 DR Pfam; PF00092; VMA; 1.  
 DR PRINTS; PRO1185; INTEGRIN.  
 DR PRINTS; PRO0453; VWFADOMAIN.  
 DR SMART; SM00327; VMA; 1.  
 DR SMART; SM00327; VMA; 1.  
 DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE; PS00242; VMA; 1.  
 DR 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Receptor; Signal;  
 KW Transmembrane.

FT SIGNAL 1 16 Integrin alpha-M.  
 FT CHAIN 17 1152  
 Query Match 14.1%; Score 132.5; DB 1; Length 1152;  
 Best Local Similarity 26.3%; Pred. No. 0.014;  
 Matches 52; Conservative 41; Mismatches 66; Indels 39; Gaps 11;

QY 3 DLVFLIDKSGSVL-HHNNIYYFVEQLAHKFIQSPMSFVSTRTGTLTKLTD-----57  
 DQ 150 DIAFLIDGSGSIIIPDRMKKEFVST-----VMEQLKSKTIFS-----LMQYSEPRRIH 199  
 QY 58 -----RQIQGLEELQKVLPGEDTVHKEFERASEQIYENRGQYRTA-SVITALT 108  
 DQ 200 FTTFKEPQNNPNPRLSLVPITQL--GRTHPATGIRKRVRELFNTNGARKNAKPLIVIT 257  
 QY 109 DGLHEDLFFYSR--REANRSRLGAIYVCVYGDVFEETOLAR-----IADS--KDHVFP 159  
 DQ 258 DGEKFGDPLGIEDYIPEADRE--GVIRYVIGGDARFSEKSRQELNTIASKPRDHFQ 314  
 QY 160 VNDGFQALGGIHSILKK 177  
 DQ 315 VNN-FBALKTITQNLREK 331

RESULT 14  
 Q8CEK9 PRELIMINARY; PRT; 1182 AA.  
 AC Q8CEK9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN full-  
 DE length enriched library, clone:E330019814 product:hypothetical von  
 DE Willebrand factor type A domain containing protein, full insert  
 DE sequence.  
 GN Name=E330026B02Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Ovary;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Ovary;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Ovary;  
 RA The RIKEN Genome Exploration, Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Ovary;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN=C57BL/6J; TISSUE=Ovary;
RA MEDLINE=20530913; PubMed=11075861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Komno H., Akiyama Y., Nishi K., Kitanai T., Tashiro H., Itoh M.,
RA Saito N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kasahagi K.,
RA Fujitake S., Inoue K., Togawa Y., Izawa M., Ohara E., Matsubiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Mizumatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN Integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Ovary;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirooka T., Hirozane T.,
RA Hori F., Imocani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Komno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK054356; BAC3749.1; -.
DR HSSP; P11215; IMF7.
DR MGD; MGI:2444259; E330026B02R1K.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00092; VMA; 6.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VMA; 6.
DR PROSITE; PS50234; VWF_A; 6.
DR Hypothetical protein.
KM SEQUENCE 1182 AA; 130946 MW; CA62P1926A426424 CRC64;
SQ
Query Match 13.7%; Score 129.5; DB 2; Length 1182;
Best Local Similarity 22.0%; Pred. No. 0.026;
Matches 42; Conservative 59; Mismatches 65; Indels 25; Gaps 9;
QY 3 DLYEILDKSGSVL-HHNNETIYFPEOLAKHP-ISP-DLRMSFIYFSGRTGLTKLME--D 57
DB 435 DLYLIDGSGSTOPTDHEMKTFPLSEVGMFNIPHKRVGAOVAVTWDLEFISTKYSN 494
QY 58 REQIQLQELQKVLPGSDT-----YMHGFERASEQIYENRQGYRTASVIALTDG 110
DB 495 KPDLGKAIEMIRQW--GGNTVTGAALNFTLKLGRKKE-----RSKYPCHLVLTNG 546
QY 111 ELHEDLFFYSEBRANRSDGAIYCYGVQDFNETQLARLADSKDHFPPVNDGQALQGI 170
DB 547 MSRSRSV-----GPAHKLRSEINIRVHAIGVKEANOTQREIAGEKRYVYVHE-FDALRNI 601
QY 171 IHSILKSGCTE 181
DB 602 RNOVVEICAE 612

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn;
RA MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;
RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
RA Nishida Y., Ohara M., Kimata K.;
RT "The complete primary structure of type XII collagen shows a chimeric
RT molecule with reiterated fibronectin type III motifs, von Willebrand
RT factor A motifs, a domain homologous to a noncollagenous region of
RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
RT site.";
RL J. Cell Biol. 115:209-221(1991).
RN [2]
RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
RP 2846-2873.
RX MEDLINE=90062079; PubMed=2584192;
RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
RT "Type XII collagen. A large multidomain molecule with partial homology
RT to type IX collagen.";
RL J. Biol. Chem. 264:19772-19778(1989).
RN [3]
RP SEQUENCE OF 2960-3076 FROM N.A.
RX MEDLINE=87317590; PubMed=3476925;
RA Gordon M.K., Gerecke D.R., Olsen B.R.;
RT "Type XII collagen: distinct extracellular matrix component discovered
RT by cDNA cloning.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).
RN [4]
RP SEQUENCE OF 1-1283 FROM N.A. (ISOFORM SHORT), AND ALTERNATIVE
RP SPLICING.
RC TISSUE=Embryo;
RX MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-P;
RA Tneub J., Tneub B.;
RT "The two splice variants of collagen XII share a common 5' end.";
RL Biochim. Biophys. Acta 1171:97-98(1992).
RN [5]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;
RA Koch M., Bornmann B., Matchison M., Hagios C., Tneub B., Chiquet M.;
RT "Large and small splice variants of collagen XII: differential
RT expression and ligand binding.";
RL J. Cell Biol. 130:1005-1014(1995).
CC -1- FUNCTION: Type XII collagen interacts with type I collagen-
CC containing fibrils, the COL1 domain could be associated with the
CC surface of the fibrils, and the COL2 and NC3 domains may be
CC localized in the perifibrillar matrix.
CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of
CC nontriple-helical sequences.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The final tissue form of collagen XII may contain
CC homotrimers of either isoform long or isoform short or any
CC combination of isoform long and isoform short. Only isoform long
CC is a proteoglycan. Isoform long has more restricted expression
CC in embryonic tissue than isoform short;
CC Name=Long;
CC IsoId=P13944-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P13944-2; Sequence=VSP_001148;
CC -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,
CC ligaments, perichondrium, and perosteum, all dense connective
CC tissues containing type I collagen.
CC -1- DOMAIN: This sequence defines five distinct domains, two triple-
CC helical domains (COL1 and COL2) and three nontriple-helical
CC domains (NC1, NC2, and NC3).
CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at
CC each end.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- PTM: O-glycosylated; glycosaminoglycan or chondroitin-sulfate type
CC (By similarity).
CC -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC interrupted helices (FACIT) family.

```

CC -1- SIMILARITY: Contains 18 fibronectin type III domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -1- SIMILARITY: Contains 4 WFA domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; D00824; BAA0701.1; -;  
 DR EMBL; X61024; AAA43358.1; -;  
 DR EMBL; M17375; AAA48718.1; -;  
 DR EMBL; J05137; AAA46635.1; -;  
 DR EMBL; X67327; CAA4744.1; -;  
 DR PIR; A40020; A40020.  
 DR HSP; P56199; I0C5.  
 DR InterPro; IPR008160; Collagen.  
 DR InterPro; IPR009885; CONA\_1like\_1ec\_g1.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003129; TSP\_N.  
 DR InterPro; IPR002035; WFA\_2.  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF00041; fn3; 17.  
 DR Pfam; PF02210; TSP\_N; 1.  
 DR Pfam; PF00097; WFA\_4.  
 DR PRINTS; PR00453; WFAADOMAIN.  
 DR SMART; SM00060; FN3; 18.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; WFA; 4.  
 DR PROSITE; PS00853; FN3; 18.  
 DR PROSITE; PS50234; WFA; 4.  
 DR Alternative splicing; Cell adhesion; Collagen;  
 DR Direct protein sequencing; Extracellular matrix; Glycoprotein;  
 DR Hydroxylation; Repeat; Signal; Structural protein.  
 FT SIGNAL 1 24 Potential.  
 FT CHAIN 25 3124 Collagen alpha 1(XIII) chain.  
 FT DOMAIN 25 112 Fibronectin type-III 1.  
 FT DOMAIN 139 311 WFA 1.  
 FT DOMAIN 332 421 Fibronectin type-III 2.  
 FT DOMAIN 439 615 WFA 2.  
 FT DOMAIN 630 718 Fibronectin type-III 3.  
 FT DOMAIN 721 809 Fibronectin type-III 4.  
 FT DOMAIN 812 902 Fibronectin type-III 5.  
 FT DOMAIN 905 993 Fibronectin type-III 6.  
 FT DOMAIN 995 1083 Fibronectin type-III 7.  
 FT DOMAIN 1086 1175 Fibronectin type-III 8.  
 FT DOMAIN 1199 1371 WFA 3.  
 FT DOMAIN 1386 1472 Fibronectin type-III 9.  
 FT DOMAIN 1474 1564 Fibronectin type-III 10.  
 FT DOMAIN 1566 1654 Fibronectin type-III 11.  
 FT DOMAIN 1655 1745 Fibronectin type-III 12.  
 FT DOMAIN 1756 1845 Fibronectin type-III 13.  
 FT DOMAIN 1847 1935 Fibronectin type-III 14.  
 FT DOMAIN 1937 2026 Fibronectin type-III 15.  
 FT DOMAIN 2028 2117 Fibronectin type-III 16.  
 FT DOMAIN 2119 2206 Fibronectin type-III 17.  
 FT DOMAIN 2210 2294 Fibronectin type-III 18.  
 FT DOMAIN 2327 2500 TSP N-terminal.  
 FT DOMAIN 2524 2716 Nonhelical region (NC2).  
 FT DOMAIN 2455 2750 Triple-helical region (COL2) with 1  
 FT DOMAIN 2751 2902 Imperfection.  
 FT DOMAIN 2903 2945 Nonhelical region (NC2).  
 FT DOMAIN 2946 3048 Triple-helical region (COL1) with 2  
 FT DOMAIN 3049 3124 Imperfections.  
 FT DOMAIN 3086 3096 Nonhelical region (NC1).  
 FT DOMAIN 3111 3123 Asp/Glu-rich (acidic).  
 FT SITE 2899 2901 Arg/Lys-rich (basic).  
 FT 2899 2901 Cell attachment site (Potential).

FT CARBOHYD 32 32 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 797 797 O-linked (Xyl... ) (chondroitin sulfate)  
 FT CARBOHYD 890 890 O-linked (Xyl... ) (chondroitin sulfate)  
 FT CARBOHYD 981 981 O-linked (Xyl... ) (chondroitin sulfate)  
 FT CARBOHYD 1006 1006 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1032 1032 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1044 1044 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1512 1512 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1767 1767 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 2210 2210 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 2273 2273 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 2532 2532 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 2683 2683 N-linked (GlcNAc... ) (Potential).  
 FT VARSPLIC 25 1188 Missing (in isoform Short).  
 FT CONFLICT 1258 1258 /FTid=VSP\_001148.  
 FT CONFLICT 1264 1264 T -> S (in Ref. 4).  
 FT CONFLICT 2759 2759 D -> E (in Ref. 4).  
 FT CONFLICT 2803 2803 P -> A (in Ref. 2).  
 FT CONFLICT 2977 2977 L -> F (in Ref. 2).  
 FT CONFLICT 3075 3076 V -> F (in Ref. 2).  
 FT CONFLICT 3076 3076 QP -> AG (in Ref. 3).  
 FT SEQUENCE 3124 AA; 340578 MW; 094285AFEF7346CF CRC64;

Query Match 13.6%; Score 128; DB 1; Length 3124;  
 Best Local Similarity 26.3%; Pred. No 0.11; Indels 16; Gaps 9;  
 Matches 49; Conservative 42; Mismatches 79;

QY 3 DLVFIIDKSGSV-LHMNEIYYFVEQLAHKF--ISPO-LRMSFIVFSTGTTLMLKT--E 56  
 DB 2327 DIVFLTDASWSISGIDNFKVKVFYFNTVGAFLDINPAGIGVSLVQYSDAQSEFKLTWFD 2386

QY 57 DREGIRGGLSELQKVLPGCDTYHMEGFRASQIY-YENRGRTASVIALTLTGELHED 115  
 DB 2387 DKQALGALGNVQ--YRGWTRGKALTFIKKYLWESGWRRCVPPVLVVTGSRQDE 2444

QY 116 LFFYSREANRSRDIAIVCVGVKDFNETOLARIAD--SKDHYFPVNDGFOALGIIHS 173  
 DB 2445 V---RKAAIVYIQHSGSVFVGVADVYNELAKIASKPSBRHVFYDD-FDAFEKIQDN 2499

QY 174 ILKKSQ 179  
 DB 2500 LVTFVC 2505

Search completed: June 13, 2005, 20:03:17  
 Job time : 61.7608 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 19:37:17 ; Search time 64.0514 Seconds  
(without alignments)  
1044.623 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_44\_216  
Perfect score: 1 DLYFILDKSSSVLHNMNEY.....HVEPVNDGQALQGIHSIL 173  
Sequence:

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_16Dec04:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	899	100.0	297	4	AAM38976 Human pol
2	899	100.0	328	7	Adi00558 Human TAN
3	899	100.0	328	7	Adm64584 Human TAN
4	899	100.0	333	3	AA801422 Human TAN
5	899	100.0	333	5	ABP54905 Human ant
6	899	100.0	333	7	Adi00534 Human TAN
7	899	100.0	333	7	Adm64568 Human von
8	899	100.0	342	7	Adi00554 Human TAN
9	899	100.0	342	7	Adm64580 Human TAN
10	899	100.0	345	7	Adi00556 Human TAN
11	899	100.0	345	7	Adm64582 Human TAN
12	899	100.0	368	5	ABP54903 Human ant
13	899	100.0	384	5	Adm64586 TANGO197
14	899	100.0	403	4	AAE01439 Human gen
15	899	100.0	403	5	ABG63874 Human alb
16	899	100.0	403	8	Adi77139 Albumin f
17	899	100.0	460	7	Adi00562 Human TAN
18	899	100.0	460	7	Adi00560 Human TAN
19	899	100.0	479	7	Adi00564 Human TAN
20	899	100.0	504	7	Adi00566 Human TAN
21	899	100.0	529	7	Adi00568 Human TAN
22	899	100.0	538	7	AdE97421 Human chi
23	899	100.0	540	7	Adi00544 Human TAN
24	899	100.0	540	7	Adm64588 TANGO197
25	899	100.0	549	7	Adi00546 Human TAN

26	899	100.0	549	7	Adi00542 Human TAN
27	899	100.0	549	7	Adi00548 Human TAN
28	899	100.0	549	7	Adm64590 TANGO197
29	899	100.0	549	7	Adm64592 TANGO197
30	899	100.0	551	7	Adi00550 Human TAN
31	899	100.0	551	7	Adm64576 Mouse TAN
32	899	100.0	564	5	ABB90750 Human Tum
33	899	100.0	564	5	ABB90724 Human Tum
34	899	100.0	564	5	ABP54904 Human ant
35	899	100.0	564	6	ABU54457 Human cum
36	899	100.0	564	6	ABU54431 Human cum
37	899	100.0	564	7	Adi00552 Human TAN
38	899	100.0	564	7	Adi70017 Human hea
39	899	100.0	564	7	Adm64578 Human TAN
40	899	100.0	564	8	AdR48216 Human cum
41	894	99.4	403	4	AAE01469 Human gen
42	894	99.4	403	5	ABG63873 Human alb
43	894	99.4	403	8	Adi77138 Albumin f
44	890	99.0	562	5	ABB90731 Mouse Tum
45	890	99.0	562	5	ABB90785 Mouse Tum

#### ALIGNMENTS

RESULT 1  
AAM38976  
ID AAM38976 standard; protein; 297 AA.  
XX  
AC AAM38976;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 2121.  
XX  
KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN W020015312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US034263.  
XX  
PR 23-DEC-1999; 99US-00471275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-00552317.  
PR 20-JUN-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 29-NOV-2000; 2000US-00727344.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Aeundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou F, Goodrich R, Drmanac RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
DR N-PSDB; AA158132.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders such  
PT as central nervous system injuries.  
XX  
PS Example 4; SEQ ID NO 2121; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the  
CC encoded polypeptides (AA38642-AA44213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukemias and  
CC C.N.S disorders. Note: The sequence data for this patent did not form  
CC part of the printed specification  
XX  
SQ Sequence 297 AA;

Query Match 100.0%; Score 899; DB 4; Length 297;  
Best Local Similarity 100.0%; Pred. No. 2,4e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 DLVFIIDKSGSVLHHNNEIYFVEQLAKRFISPOLRMSFIVFSTRGTTLMKLTEDRQIR 60  
|||  
Db 44 DLVFIIDKSGSVLHHNNEIYFVEQLAKRFISPOLRMSFIVFSTRGTTLMKLTEDRQIR 103  
  
Qy 61 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTDEGELHEDLFFYSER 120  
|||  
Db 104 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTDEGELHEDLFFYSER 163  
  
Qy 121 EANRSRLGAIIVYCVGKDFNETQRLARIDSKDHVPVNDGFQALQGIHSIL 173  
|||  
Db 164 EANRSRLGAIIVYCVGKDFNETQRLARIDSKDHVPVNDGFQALQGIHSIL 216

RESULT 2  
AD100558  
ID AD100558 standard; protein; 328 AA.

XX AD100558;

XX 22-Apr-2004 (first entry)

XX Human TANGO 197 HisTag fusion protein - plasmid pO615.

XX fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KM cutaneous; inhalation anthrax; human; TANGO 197 HisTag fusion; mutant;  
KM plasmid pO615; mutcin.

XX Homo sapiens.

OS Synthetic.

XX US2003144193-A1.

XX 31-JUL-2003.

XX 24-JUL-2002; 2002US-00201292.

XX 20-DEC-2001; 2001US-00038307.

XX (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.

PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

DR WPI; 2003-720708/68.

DR N-PSDB; AD100557.

XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like

PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.

XX Claim 45; SEQ ID NO 26; 86pp; English.

XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 HisTag fusion protein of the invention.  
XX

SQ Sequence 328 AA;

Query Match 100.0%; Score 899; DB 7; Length 328;  
Best Local Similarity 100.0%; Pred. No. 2,7e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 DLVFIIDKSGSVLHHNNEIYFVEQLAKRFISPOLRMSFIVFSTRGTTLMKLTEDRQIR 60  
|||  
Db 44 DLVFIIDKSGSVLHHNNEIYFVEQLAKRFISPOLRMSFIVFSTRGTTLMKLTEDRQIR 103  
  
Qy 61 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTDEGELHEDLFFYSER 120  
|||  
Db 104 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTDEGELHEDLFFYSER 163  
  
Qy 121 EANRSRLGAIIVYCVGKDFNETQRLARIDSKDHVPVNDGFQALQGIHSIL 173  
|||  
Db 164 EANRSRLGAIIVYCVGKDFNETQRLARIDSKDHVPVNDGFQALQGIHSIL 216

RESULT 3  
ADM64584  
ID ADM64584 standard; protein; 328 AA.

XX ADM64584;

XX 03-JUN-2004 (first entry)

XX Human TANGO197-His tag fusion protein #2.

XX antibacterial; gene therapy;  
KM von Willebrand factor A-like domain amino acid sequence;  
KM vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
KM inhalation anthrax; human; TANGO197; his tag; fusion protein.

XX Homo sapiens.

OS Synthetic.

XX US2003134786-A1.

XX 17-JUL-2003.

XX 20-DEC-2001; 2001US-00038307.

XX 20-DEC-2001; 2001US-00038307.

XX (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.

PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;

DR WPI; 2003-829643/77.

DR N-PSDB; ADM64583.

XX New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.



PS Claim 44; SEQ ID NO 26; 64pp; English.

XX The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (VWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the VWF. Also described are: a method of preventing or  
 CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to Bacillus anthracis;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
 CC sequence of a fusion protein comprising mature human TANGO197, thrombin  
 CC cleavage site and his tag that can be used to treat exposure to or  
 CC prevent a symptom of anthrax.

XX  
 SQ Sequence 328 AA;

Query Match 100.0%; Score 899; DB 7; Length 328;  
 Best Local Similarity 100.0%; Pred. No. 2,7e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHMHNEIYFVEQLAHKFTSPOLRMSFIVFSTRTTLMKLTEDREQIR 60  
 DB 44 DLYFILDKSGSVLHMHNEIYFVEQLAHKFTSPOLRMSFIVFSTRTTLMKLTEDREQIR 103

QY 61 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTGELHEDLFFYSER 120  
 DB 104 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTGELHEDLFFYSER 163

QY 121 EANSRDLGAIIVCVGVDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 173  
 DB 164 EANSRDLGAIIVCVGVDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 216

RESULT 4  
 ID AAB01422 standard; protein; 333 AA.

AC AAB01422;  
 DT 20-OCT-2000 (first entry)  
 DE Human TANGO 197.

XX TANGO; 128; 140; 197; 212; 213; 224; 239; modulating agent; asthma;  
 KM graft versus-host diseases; rheumatoid arthritis; psoriasis;  
 KM inflammatory bowel disease; septic shock; ulcerative colitis;  
 KM Crohn's disease; chronic myelogenous leukemia; cancer; liver disease;  
 KM Hodgkin's disease; osteoarthritis; Lyme's disease; cachexia;  
 KM autoimmune disease; myasthenia gravis; autoimmune diabetes;  
 KM systemic lupus erythematosus; transgenic animal; diagnosis; prognosis;  
 KM prophylactic; therapeutic; human.

XX Homo sapiens.  
 OS  
 XX  
 PN WO200039284-A1.  
 PD 06-JUL-2000.  
 PF 23-DEC-1999; 99WO-US031025.  
 XX  
 PR 30-DEC-1998; 98US-00223546.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Holtzman DA;  
 XX  
 DR WPI; 2000-465743/40.  
 DR N-PSDB; AAA47455.  
 XX  
 PT Novel nucleic acid sequences encoding TANGO-128, 140, 197, 212, 213, 224  
 PT and 239 polypeptides useful for the treatment of asthma, rheumatoid  
 PT arthritis, psoriasis and autoimmune diseases.  
 XX

PS Claim 8; Fig 4; 209pp; English.

XX Nucleic acids encoding TANGO polypeptides are useful as modulating agents  
 CC for regulating cellular processes like asthma, graft versus-host  
 CC diseases, rheumatoid arthritis, psoriasis, inflammatory bowel disease,  
 CC septic shock, ulcerative colitis, Crohn's disease, chronic myelogenous  
 CC leukemia, cancer, liver disease, Hodgkin's disease, osteoarthritis,  
 CC Lyme's disease, cachexia and autoimmune diseases e.g. myasthenia gravis,  
 CC autoimmune diabetes and systemic lupus erythematosus. The nucleic acids  
 CC are also useful for producing transgenic animals and the TANGO  
 CC polypeptides themselves. Partial TANGO-128, 140, 197, 212, 213, 224, 239  
 CC sequences are useful in forensic biology, for diagnostic assays,  
 CC prognostic assays, pharmacogenomics and for monitoring clinical trials.  
 CC TANGO polypeptides are suitable for both prophylactic and therapeutic  
 CC methods for treating a subject at risk of a disorder or having a disorder  
 CC associated with aberrant TANGO expression. A wide range of cellular  
 CC disorders can be treated

XX  
 SQ Sequence 333 AA;

Query Match 100.0%; Score 899; DB 3; Length 333;  
 Best Local Similarity 100.0%; Pred. No. 2,8e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHMHNEIYFVEQLAHKFTSPOLRMSFIVFSTRTTLMKLTEDREQIR 60  
 DB 44 DLYFILDKSGSVLHMHNEIYFVEQLAHKFTSPOLRMSFIVFSTRTTLMKLTEDREQIR 103

QY 61 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTGELHEDLFFYSER 120  
 DB 104 QGLEELQKVLPGSDTYMHGEFERASEQIYYENRQGYRTASVITALTGELHEDLFFYSER 163

QY 121 EANSRDLGAIIVCVGVDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 173  
 DB 164 EANSRDLGAIIVCVGVDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 216

RESULT 5  
 ID ABP54905 standard; protein; 333 AA.

AC ABP54905;  
 DT 08-JAN-2003 (first entry)  
 DE Human anthrax toxin receptor.

XX Anthrax; toxin; receptor; human; antibacterial.  
 KM  
 KM Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT /label= Signal\_peptide  
 FT Protein 28..333  
 FT /label= Mature\_protein

XX WO200246228-A2.  
 PN  
 PD 13-JUN-2002.  
 PF 03-OCT-2001; 2001WO-US030941.  
 XX  
 PR 05-DEC-2000; 2000US-0251481P.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Young JAT, Bradley KA, Collier RJ, Mogridge JS;  
 XX  
 DR WPI; 2002-713235/77.  
 DR N-PSDB; ABV73883.  
 XX  
 PT Novel isolated polypeptide useful for identifying agent that prevents or

PT reduces effect of anthrax toxin on host cell, for treating human or non-  
PT human animal suffering from anthrax.  
XX  
PS Claim 1, Page 42-43; 45pp; English.  
XX  
CC The present sequence is the protein sequence of a polypeptide identified  
CC as a human anthrax toxin receptor (ATR) polypeptide on the basis of  
CC identity to a newly isolated human ATR (see ABP54903). The 2 polypeptides  
CC are identical between amino acids 1-317, but differ thereafter at the C-  
CC terminus. The present polypeptide, encoded by IMAGE CLONE 4563020, has no  
CC previously known function, and there has been no prior indication that it  
CC is a complete or partial ATR. The invention provides ATR polypeptides and  
CC polynucleotides, vectors, host cells, and transgenic and knock-out  
CC animals. It also provides methods for identifying molecules that bind the  
CC ATR and which reduce the toxicity of anthrax toxin. A claimed method for  
CC treating anthrax in a human or animal involves administering an agent  
CC that inhibits binding between anthrax toxin protective antigen (PA) and  
CC ATR at a level effective to reduce the severity of anthrax. Suitable  
CC agents include the present polypeptide or a PA-binding fragment of it, a  
CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
CC nucleic acid  
XX  
SQ Sequence 333 AA;  
XX  
Query Match 100.0%; Score 899; DB 5; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2.8e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLVFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFIVFSRGTTLMLKLTEDRQIR 60  
DB 44 DLVFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFIVFSRGTTLMLKLTEDRQIR 103  
QY 61 QGSEELQKVLPGDPTVHMEGFERRASEQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGSEELQKVLPGDPTVHMEGFERRASEQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 173  
DB 164 EANRSRLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 216  
RESULT 6  
ADI00534  
ID ADI00534 standard; protein; 333 AA.  
XX  
AC ADI00534;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human TANGO 197 protein.  
XX  
KM fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
XX cutaneous; inhalation anthrax; human; TANGO 197.  
XX  
OS Homo sapiens.  
XX  
PN US2003144193-A1.  
XX  
PD 31-JUL-2003.  
XX  
PF 24-JUL-2002; 2002US-00201292.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-720708/68.

DR N-PSDB; ADI00533.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
PS Claim 26; SEQ ID NO 2; 86pp; English.  
XX  
XX The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 protein of the invention.  
XX  
SQ Sequence 333 AA;  
XX  
Query Match 100.0%; Score 899; DB 7; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2.8e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLVFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFIVFSRGTTLMLKLTEDRQIR 60  
DB 44 DLVFIIDKSGSVLHHNMEIYFVEQLAHKFIISPOLRMSFIVFSRGTTLMLKLTEDRQIR 103  
QY 61 QGSEELQKVLPGDPTVHMEGFERRASEQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGSEELQKVLPGDPTVHMEGFERRASEQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 173  
DB 164 EANRSRLGAIYVCVGVKDFNETQLARIADSKDHVPVNDGFQALQGIHSIL 216  
RESULT 7  
ADM64568  
ID ADM64568 standard; protein; 333 AA.  
XX  
AC ADM64568;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human von Willebrand factor A-like domain protein TANGO197.  
XX  
KM antibacterial; gene therapy;  
KM von Willebrand factor A-like domain amino acid sequence;  
KM vWF amino acid sequence; anthrax; Bacillus anthracis; cutaneous anthrax;  
XX inhalation anthrax; human; TANGO197.  
XX  
OS Homo sapiens.  
XX  
PN US2003134786-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 20-DEC-2001; 2001US-00038307.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-829643/77.  
XX  
DR N-PSDB; ADM64567.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like

PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
PS Claim 26; SEQ ID NO 2; 64pp; English.  
XX  
CC The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to *Bacillus anthracis*;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
CC sequence of a human von Willebrand factor A-like domain (vWF) amino acid  
CC sequence TANGO197.  
XX  
SQ Sequence 333 AA;  
XX  
Query Match 100.0%; Score 899; DB 7; Length 333;  
Best Local Similarity 100.0%; Pred. No. 2.8e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLVFIIDKSGSVLHNMNIYFVQLAHKFTSPQLRMSFVFSRGTTLTKLTEDRQIR 60  
DB 44 DLVFIIDKSGSVLHNMNIYFVQLAHKFTSPQLRMSFVFSRGTTLTKLTEDRQIR 103  
QY 61 QGLEBELQKVLPGGDTVMHGEFPERASEQIYYENRGYRTASVITALTGELHEDLFFYSER 120  
DB 104 QGLEBELQKVLPGGDTVMHGEFPERASEQIYYENRGYRTASVITALTGELHEDLFFYSER 163  
QY 121 EANRSRDLGAIIVYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 173  
DB 164 EANRSRDLGAIIVYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 216  
XX  
RESULT 8  
AD100554  
ID AD100554 standard; protein; 342 AA.  
XX  
AC AD100554;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human TANGO 197 FLAG fusion protein - plasmid p0613.  
XX  
KW fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KW cutaneous; inhalation anthrax; human; TANGO 197 FLAG fusion; mutant;  
KW plasmid p0613; mutain.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN US2003144193-A1.  
XX  
PD 31-JUL-2003.  
XX  
PF 24-JUL-2002; 2002US-00201292.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-720708/68.  
XX  
DR N-PSDB; AD100553.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.

PT to the vWF.  
XX  
PS Claim 45; SEQ ID NO 22; 86pp; English.  
XX  
CC The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 FLAG fusion protein of the invention.  
XX  
SQ Sequence 342 AA;  
XX  
Query Match 100.0%; Score 899; DB 7; Length 342;  
Best Local Similarity 100.0%; Pred. No. 2.9e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DLVFIIDKSGSVLHNMNIYFVQLAHKFTSPQLRMSFVFSRGTTLTKLTEDRQIR 60  
DB 53 DLVFIIDKSGSVLHNMNIYFVQLAHKFTSPQLRMSFVFSRGTTLTKLTEDRQIR 112  
QY 61 QGLEBELQKVLPGGDTVMHGEFPERASEQIYYENRGYRTASVITALTGELHEDLFFYSER 120  
DB 113 QGLEBELQKVLPGGDTVMHGEFPERASEQIYYENRGYRTASVITALTGELHEDLFFYSER 172  
QY 121 EANRSRDLGAIIVYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 173  
DB 173 EANRSRDLGAIIVYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 225  
XX  
RESULT 9  
ADM64580  
ID ADM64580 standard; protein; 342 AA.  
XX  
AC ADM64580;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human TANGO197-FLAG epitope fusion protein.  
XX  
KW antibacterial; gene therapy;  
KW von Willebrand factor A-like domain amino acid sequence;  
KW vWF amino acid sequence; anthrax; *Bacillus anthracis*; cutaneous anthrax;  
KW inhalation anthrax; human; TANGO197; FLAG; fusion protein; gene.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN US2003134786-A1.  
XX  
PD 17-JUL-2003.  
XX  
PF 20-DEC-2001; 2001US-00038307.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-829643/77.  
XX  
DR N-PSDB; ADM64579.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX  
PS Claim 44; SEQ ID NO 22; 64pp; English.

XX The invention describes a fusion polypeptide comprising a von Willebrand  
CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
CC heterologous to the vWF. Also described are: a method of preventing or  
CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
CC exposure to or suspected of having been exposed to *Bacillus anthracis*;  
CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
CC The composition and method are useful in preventing or ameliorating  
CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
CC sequence of a fusion protein comprising human TANGO197 and FLAG that can  
CC be used to treat exposure to or prevent a symptom of anthrax.  
XX  
SQ Sequence 342 AA;  
  
Query Match 100.0%; Score 899; DB 7; Length 342;  
Best Local Similarity 100.0%; Pred. No. 2,9e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLVFIIDKSGSVLHHNNEIYFVEQLAKHFIISPOLRMSFIVFSTRGTTLMKLTEDREQIR 60  
Db 53 DLVFIIDKSGSVLHHNNEIYFVEQLAKHFIISPOLRMSFIVFSTRGTTLMKLTEDREQIR 112  
  
QY 61 QGLEBLQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFYSER 120  
Db 113 QGLEBLQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFYSER 172  
  
QY 121 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHSIL 173  
Db 173 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHSIL 225  
  
RESULT 10  
ADM64556  
ID ADM64556 standard; protein; 345 AA.  
XX  
AC ADI00556;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human TANGO 197 HisTag fusion protein - plasmid p0614.  
XX  
KM fusion; von Willebrand factor A-like domain; vWF; antibacterial;  
KM cutaneous; inhalation anthrax; human; TANGO 197 HisTag fusion; mutant;  
KM plasmid p0614; mutain.  
XX  
OS Homo sapiens.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 331 /note= "Wild-type Cys replaced by Ser followed by  
FT thrombin cleavage site and His tag"  
XX  
PN US2003144193-A1.  
XX  
PD 31-JUL-2003.  
XX  
PR 24-JUL-2002; 2002US-00201292.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-720708/68.  
XX  
DR N-PSDB; ADI00555.  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like

PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
XX to the vWF.  
XX  
PS Claim 45; SEQ ID NO 24; 86pp; English.  
XX  
XX  
CC The invention relates to a novel fusion polypeptide comprising a von  
CC Willebrand factor A-like domain (vWF) amino acid sequence and an amino  
CC acid sequence heterologous to the vWF. The polypeptide of the invention  
CC demonstrates antibacterial activities whilst the composition and method  
CC may be useful in preventing or ameliorating the symptoms of cutaneous  
CC and/or inhalation anthrax. The current sequence is that of the human  
CC TANGO 197 HisTag fusion protein of the invention.  
XX  
SQ Sequence 345 AA;  
  
Query Match 100.0%; Score 899; DB 7; Length 345;  
Best Local Similarity 100.0%; Pred. No. 2,9e-92;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DLVFIIDKSGSVLHHNNEIYFVEQLAKHFIISPOLRMSFIVFSTRGTTLMKLTEDREQIR 60  
Db 44 DLVFIIDKSGSVLHHNNEIYFVEQLAKHFIISPOLRMSFIVFSTRGTTLMKLTEDREQIR 103  
  
QY 61 QGLEBLQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFYSER 120  
Db 104 QGLEBLQKVLPGGDTYMHGFERASEQIYYENRQGYRTASVITALTDEGLHEDLFFYSER 163  
  
QY 121 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHSIL 173  
Db 164 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHSIL 216  
  
RESULT 11  
ADM64582  
ID ADM64582 standard; protein; 345 AA.  
XX  
AC ADM64582;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE Human TANGO197-His tag fusion protein #1.  
XX  
KM antibacterial; gene therapy;  
KM von Willebrand factor A-like domain amino acid sequence;  
KM vWF amino acid sequence; anthrax; *Bacillus anthracis*; cutaneous anthrax;  
KM inhalation anthrax; human; TANGO197; his tag; fusion protein; gene.  
XX  
OS Homo sapiens.  
XX  
OS Synthetic.  
XX  
PN US2003134786-A1.  
XX  
PD 17-JUL-2003.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PR 20-DEC-2001; 2001US-00038307.  
XX  
PA (ROTT/) ROTTMAN J B.  
PA (OKEE/) O'KEEFE T L.  
PA (OZKA/) OZKAYNAK E.  
PA (HEAL/) HEALEY J J.  
XX  
PI Rottman JB, O'Keefe TL, Ozkaynak E, Healey JJ;  
XX  
DR WPI; 2003-829643/77.  
XX  
DR N-PSDB; ADM64581.  
XX  
XX  
PT New TANGO 197 and/or TANGO 216 fusion polypeptides for preventing or  
PT ameliorating symptoms of anthrax comprises a von Willebrand factor A-like  
PT domain (vWF) amino acid sequence and an amino acid sequence heterologous  
PT to the vWF.  
XX

PS Claim 44, SEQ ID NO 24; 64pp; English.

CC The invention describes a fusion polypeptide comprising a von Willebrand  
 CC factor A-like domain (vWF) amino acid sequence and an amino acid sequence  
 CC heterologous to the vWF. Also described are: a method of preventing or  
 CC ameliorating a symptom of anthrax in a subject thought to be at risk for  
 CC exposure to or suspected of having been exposed to Bacillus anthracis;  
 CC and a pharmaceutical composition comprising the novel fusion polypeptide.  
 CC The composition and method are useful in preventing or ameliorating  
 CC symptoms of cutaneous and/or inhalation anthrax. This is the amino acid  
 CC sequence of a fusion protein comprising human TANGO197, thrombin cleavage  
 CC site and his tag that can be used to treat exposure to or prevent a  
 CC symptom of anthrax.

XX Sequence 345 AA;

Query Match 100.0%; Score 899; DB 7; Length 345;  
 Best Local Similarity 100.0%; Pred. No. 2,9e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNMNIYFVYEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDRQIR 60  
 DB 44 DLYFIIDKSGSVLHNMNIYFVYEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDRQIR 103  
 QY 61 OGLEBELQKVLPGDDTYHGEFPERASQIYYENRQGYRTASVITIALTDGELHEDLFFYSER 120  
 DB 104 OGLEBELQKVLPGDDTYHGEFPERASQIYYENRQGYRTASVITIALTDGELHEDLFFYSER 163  
 QY 121 EANRSRDLGAIYCVGVKDFENETOLARIADSKDHVPVNDGFQALQGIHISIL 173  
 DB 164 EANRSRDLGAIYCVGVKDFENETOLARIADSKDHVPVNDGFQALQGIHISIL 216

RESULT 12  
 ID ABP54903 standard; protein: 368 AA.  
 XX  
 AC ABP54903;  
 XX  
 DT 08-JAN-2003 (first entry)  
 XX  
 DE Human anthrax toxin receptor.  
 KW Anthrax; toxin; receptor; human; antibacterial.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..27  
 FT Region /label= signal\_peptide  
 FT 27..321  
 FT /note= "PA-binding fragment, begins at any amino acid in  
 FT the range 27-43 and ends at any amino acid in the range  
 FT 221-321, region specifically described in Claim 3"  
 FT 28..368  
 FT /label= Mature\_protein  
 FT 28..320  
 FT /note= "extracellular domain"  
 FT 44..216  
 FT /note= "von Willebrand factor A domain"  
 FT 50  
 FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
 FT motif with amino acid residues 52, 54, 118 and 150"  
 FT 52  
 FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
 FT motif with amino acid residues 50, 54, 118 and 150"  
 FT 54  
 FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
 FT motif with amino acid residues 50, 52, 118 and 150"  
 FT 118  
 FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
 FT motif with amino acid residues 50, 52, 54 and 150"  
 FT 150  
 FT Region

FT /note= "forms metal ion-dependent adhesion site (MIDAS)  
 FT motif with amino acid residues 50, 52, 54 and 118"  
 FT 320..343  
 FT /note= "putative transmembrane domain"  
 FT 344..368  
 FT /note= "cytoplasmic domain"  
 PN WO200246228-A2.  
 PD 13-JUN-2002.  
 XX  
 PF 03-OCT-2001; 2001MO-US030941.  
 XX  
 PR 05-DEC-2000; 2000US-0251481P.  
 XX  
 PA (WISC ) WISCONSIN ALUMNI RES FOUND.  
 PI Young JAT, Bradley KA, Collier RJ, Mogridge JS,  
 XX  
 DR WPI; 2002-713235/77.  
 DR N-PSDB; ABV73881.  
 XX  
 PT Novel isolated polypeptide useful for identifying agent that prevents or  
 PT reduces effect of anthrax toxin on host cell, for treating human or non-  
 PT human animal suffering from anthrax.  
 XX  
 PS Claim 1; Page 29-30; 45pp; English.

CC The present sequence is the protein sequence of a human surface- bound  
 CC anthrax toxin receptor (ATR), as predicted from an isolated cDNA clone.  
 CC Anthrax toxin protective antigen (PA) binds to the ATR at a von  
 CC Willebrand factor A domain located in the extracellular domain of ATR.  
 CC The invention provides ATR polypeptides and polynucleotides, vectors,  
 CC host cells, and transgenic and knock-out animals. It also provides  
 CC methods for identifying molecules that bind the ATR and which reduce the  
 CC toxicity of anthrax toxin. A claimed method for treating anthrax in a  
 CC human or animal involves administering an agent that inhibits binding  
 CC between PA and ATR at a level effective to reduce the severity of  
 CC anthrax. Suitable agents include ATR or a PA-binding fragment of ATR, a  
 CC PA-binding polypeptide at least 80% identical to these, a fusion protein,  
 CC a monoclonal or polyclonal antibody, a polysaccharide, a lipid or a  
 CC nucleic acid  
 XX  
 SQ Sequence 368 AA;

Query Match 100.0%; Score 899; DB 5; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 3,2e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNMNIYFVYEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDRQIR 60  
 DB 44 DLYFIIDKSGSVLHNMNIYFVYEQLAHKFISPOLRMSFIVFSTRTTLMKLTEDRQIR 103  
 QY 61 OGLEBELQKVLPGDDTYHGEFPERASQIYYENRQGYRTASVITIALTDGELHEDLFFYSER 120  
 DB 104 OGLEBELQKVLPGDDTYHGEFPERASQIYYENRQGYRTASVITIALTDGELHEDLFFYSER 163  
 QY 121 EANRSRDLGAIYCVGVKDFENETOLARIADSKDHVPVNDGFQALQGIHISIL 173  
 DB 164 EANRSRDLGAIYCVGVKDFENETOLARIADSKDHVPVNDGFQALQGIHISIL 216

RESULT 13  
 ID ADM64586 standard; protein: 384 AA.  
 XX  
 AC ADM64586;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE TANGO197 extracellular domain-mutant IGG Fc fusion protein #1.  
 XX  
 KW antibacterial; gene therapy;



80 Sequence 403 AA;  
 Query Match 100.0%; Score 899; DB 4; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFTSPQLRMSFIVFSTRTTLMKLTEDREQIR 60  
 DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFTSPQLRMSFIVFSTRTTLMKLTEDREQIR 103

QY 61 QGLEELQKVLPGGDTYMHGPERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
 DB 104 QGLEELQKVLPGGDTYMHGPERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163

QY 121 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHISIL 173  
 DB 164 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHISIL 216

RESULT 15  
 ABG63874  
 ID ABG63874 standard; protein; 403 AA.  
 AC ABG63874;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Human albumin fusion protein #549.  
 XX  
 KM Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KM human serum albumin; HSA; cancer; reproductive disorder;  
 KM digestive disorder; immune disorder; endocrine disorder;  
 KM hematopoietic disorder; neural disorder; connective disorder;  
 KM cytostatic; antifertility; antiinflammatory; antiulcer;  
 KM immunomodulator; anti-HIV; antidiabetic; haemostatic; noctropic;  
 KM neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KM osteopathic; antirheumatic.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS  
 PN W0200177137-A1.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 12-APR-2001; 2001WO-US011988.  
 XX  
 PR 12-APR-2000; 2000US-0229358P.  
 PR 25-APR-2000; 2000US-0199384P.  
 PR 21-DEC-2000; 2000US-0256931P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Haseltine WA;  
 XX  
 DR WPI; 2002-010886/01.  
 XX  
 PT New fusion protein for treating disease e.g. diabetes comprises an  
 albumin fused to a therapeutic protein.  
 XX  
 PS Claim 1; Page 874-875; 2102PP; English.  
 XX  
 CC The present invention relates to albumin fusion proteins comprising a  
 therapeutic protein X and human albumin (HA, also known as human serum  
 albumin, HSA). The proteins are useful for treating a disease or disorder  
 that may be modulated by therapeutic protein X. The albumin extends the  
 shelf-life of protein X, and may increase its biological in vitro/in vivo  
 activity. The protein is useful for treating and diagnosing disorders  
 such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's  
 disease, ulcerative colitis), immune disorders (e.g. acquired  
 immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),  
 hematopoietic disorders, neural disorders (e.g. Alzheimer's,  
 Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,

CC schizizophrenia), and connective disorders (e.g. osteoporosis, arthritis).  
 CC ABG63326-ABG65518 represent albumin fusion proteins of the invention  
 XX  
 SO Sequence 403 AA;  
 Query Match 100.0%; Score 899; DB 5; Length 403;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-92;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFTSPQLRMSFIVFSTRTTLMKLTEDREQIR 60  
 DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFTSPQLRMSFIVFSTRTTLMKLTEDREQIR 103

QY 61 QGLEELQKVLPGGDTYMHGPERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
 DB 104 QGLEELQKVLPGGDTYMHGPERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163

QY 121 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHISIL 173  
 DB 164 EANRSRLGAIIVYCVGVKDFNETOLARIADSKDHVPVNDGFQALQGIHISIL 216

Search completed: June 13, 2005, 19:56:53  
 Job time : 65.0514 secs

**THIS PAGE BLANK (USPTO)**





Db 142 DIAFLIDSGSINORDPAQMDPVKALMGEPASTSTLPSLMQYSNLKHTFTTEFNIL 201  
Qy 55 DREQIRQGLEBELQKVLPGSDTYMHEGPERASEQIYYENRQGYRTA-SVIALTDTGELHED 113  
Db 202 DQSLVDPIVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKKILLVITDGQKYRD 256  
Qy 114 LFFYSE--REANRSRDIGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 164  
Db 257 PLEYSDVTPAADKA---GIIRYALGVGDAPQEFALKEINTIGSAPPQDHVFKGN-FAA 312  
Qy 165 LQGI 168  
Db 313 LRSI 316

RESULT 2  
US-08-485-618-37  
Sequence 37, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-485-618-37

Query Match 15.4%; Score 138.5; DB 1; Length 1151;  
Best Local Similarity 28.3%; Pred. No. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;  
Qy 1 DLVFLIDKSGSV-LHNMNRIYFVEQIAHKFISPOLRMSPIVSTRTGTLMLKLT-----54  
Db 142 DIAFLIDSGSINORDPAQMDPVKALMGEPASTSTLPSLMQYSNLKHTFTTEFNIL 201

Qy 55 DREQIRQGLEBELQKVLPGSDTYMHEGPERASEQIYYENRQGYRTA-SVIALTDTGELHED 113  
Db 202 DQSLVDPIVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKKILLVITDGQKYRD 256  
Qy 114 LFFYSE--REANRSRDIGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 164  
Db 257 PLEYSDVTPAADKA---GIIRYALGVGDAPQEFALKEINTIGSAPPQDHVFKGN-FAA 312  
Qy 165 LQGI 168  
Db 313 LRSI 316

RESULT 3  
US-08-362-652-37  
Sequence 37, Application US/08362652  
Patent No. 5766850  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/362,652  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32391  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-652-37

Query Match 15.4%; Score 138.5; DB 1; Length 1151;  
Best Local Similarity 28.3%; Pred. No. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;  
Qy 1 DLVFLIDKSGSV-LHNMNRIYFVEQIAHKFISPOLRMSPIVSTRTGTLMLKLT-----54  
Db 142 DIAFLIDSGSINORDPAQMDPVKALMGEPASTSTLPSLMQYSNLKHTFTTEFNIL 201  
Qy 55 DREQIRQGLEBELQKVLPGSDTYMHEGPERASEQIYYENRQGYRTA-SVIALTDTGELHED 113  
Db 202 DQSLVDPIVQLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKKILLVITDGQKYRD 256  
Qy 114 LFFYSE--REANRSRDIGAIYVCVGD-FNE-TQLARI-----ADSKDHVPVNDGFOA 164

Db 257 PLEYSVDIIPADKA--GIIRYAIAGVDARFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 312  
QY 165 LOGI 168  
Db 313 LRSI 316

## RESULT 4

US-08-605-672-37  
Sequence 37, Application US/08605672  
Patent No. 5817515  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seer Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/605,672  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-605-672-37

Query Match 15.4%; Score 138.5; DB 2; Length 1151;  
Best Local Similarity 28.3%; Pred. No. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;  
QY 1 DLVFIIDKSGSV-LHMHNEIYYFVEQLAHKRFISPOLMSFIVPSTRTGTLMLKLE----- 54  
Db 142 DIAFLIGSSSINORDPAQKMDFFKALMGEPASTSTLFSLMQYSNLIKHTFTFEPKNIL 201  
QY 55 DREQIRGSELQKVLPGSGDTYMHGEPERASEQIYYENRQGYRTA-SVIALTLTGELHED 113  
Db 202 DPOSIVDPYIQLQ-----GLTYATGIRTYMELFHSKNGSRKSAKKILLVITDGQRYRD 256  
QY 114 LFPYSE--REANSRDLGALIVCYGVKD-FNE-TQLARI-----ADSKDHVPVNDGFOA 164  
Db 257 PLEYSVDIIPADKA--GIIRYAIAGVDARFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 312

QY 165 LOGI 168  
Db 313 LRSI 316

## RESULT 5

US-08-482-293A-37  
Sequence 37, Application US/08482293A  
Patent No. 5831029  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seer Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32684  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-482-293A-37

Query Match 15.4%; Score 138.5; DB 2; Length 1151;  
Best Local Similarity 28.3%; Pred. No. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;  
QY 1 DLVFIIDKSGSV-LHMHNEIYYFVEQLAHKRFISPOLMSFIVPSTRTGTLMLKLE----- 54  
Db 142 DIAFLIGSSSINORDPAQKMDFFKALMGEPASTSTLFSLMQYSNLIKHTFTFEPKNIL 201  
QY 55 DREQIRGSELQKVLPGSGDTYMHGEPERASEQIYYENRQGYRTA-SVIALTLTGELHED 113  
Db 202 DPOSIVDPYIQLQ-----GLTYATGIRTYMELFHSKNGSRKSAKKILLVITDGQRYRD 256  
QY 114 LFPYSE--REANSRDLGALIVCYGVKD-FNE-TQLARI-----ADSKDHVPVNDGFOA 164  
Db 257 PLEYSVDIIPADKA--GIIRYAIAGVDARFOEPALKEMLNTIGSAPPQDHVFKVGN-FAA 312  
QY 165 LOGI 168



NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 37  
LENGTH: 1151  
TYPE: PRT  
ORGANISM: Rattus rattus  
FEATURES:  
NAME/KEY: misc\_feature  
LOCATION: 361  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 464  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 506  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1117  
OTHER INFORMATION: Xaa = any or unknown amino acid  
NAME/KEY: misc\_feature  
LOCATION: 1118  
OTHER INFORMATION: Xaa = any or unknown amino acid  
US-09-688-307A-37

Query Match 15.4%; Score 138.5; DB 4; Length 1151;  
Best Local Similarity 28.3%; Pred. NO. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;

QY 1 DLYFIIDKSGSV-LHMHNEIYYFVEQLAHKFIISPOLMSFVIFSTRTTLMKLE----- 54  
DB 142 DIAFLIDGSGSINGRDPQAKMDPVKALMGEPASTLPSLMQYSNLIKHTFTFEPKNIL 201  
QY 55 DREQIROGSELQKVLFGSDTYMHEGPERASEQIYYENRQGYRTA-SVITALTDEGLHED 113  
DB 202 DPQSLVDPIVOLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKKILLVITDGQKXRD 256  
QY 114 LFPYSE--REANSRDLGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPYNDGFOA 164  
DB 257 PLEYSVDITPADKX--GIIRYALGVGDAPQEPALKELTIGSAPPQDHVKVGN-FAA 312  
QY 165 LQGI 168  
DB 313 LRGI 316

RESULT 9  
US-09-350-259-37  
Sequence 37, Application US/09350259  
Patent No. 6620915  
GENERAL INFORMATION:  
APPLICANT: Gallatin, Michael W.  
TITLE OF INVENTION: No. 6620915el Human 2  
FILE REFERENCE: 27866/35004  
CURRENT APPLICATION NUMBER: US/09/350,259  
CURRENT FILING DATE: 1999-07-08  
EARLIER APPLICATION NUMBER: 09/193,043  
EARLIER FILING DATE: 1998-11-16  
EARLIER APPLICATION NUMBER: 08/173,497  
EARLIER FILING DATE: 1993-12-23  
EARLIER APPLICATION NUMBER: 08/286,889  
EARLIER FILING DATE: 1994-08-05  
EARLIER APPLICATION NUMBER: 08/362,652  
EARLIER FILING DATE: 1994-12-21  
EARLIER APPLICATION NUMBER: 08/943,363  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 37  
LENGTH: 1151

TYPE: PRT  
ORGANISM: Rattus rattus  
US-09-350-259-37

Query Match 15.4%; Score 138.5; DB 4; Length 1151;  
Best Local Similarity 28.3%; Pred. NO. 6.6e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;

QY 1 DLYFIIDKSGSV-LHMHNEIYYFVEQLAHKFIISPOLMSFVIFSTRTTLMKLE----- 54  
DB 142 DIAFLIDGSGSINGRDPQAKMDPVKALMGEPASTLPSLMQYSNLIKHTFTFEPKNIL 201  
QY 55 DREQIROGSELQKVLFGSDTYMHEGPERASEQIYYENRQGYRTA-SVITALTDEGLHED 113  
DB 202 DPQSLVDPIVOLQ-----GLTYTATGIRTYMELFHSKNGSRKSAKKILLVITDGQKXRD 256  
QY 114 LFPYSE--REANSRDLGAIVYCVGVD-FNE-TQLARI-----ADSKDHVPYNDGFOA 164  
DB 257 PLEYSVDITPADKX--GIIRYALGVGDAPQEPALKELTIGSAPPQDHVKVGN-FAA 312  
QY 165 LQGI 168  
DB 313 LRGI 316

RESULT 10  
US-08-485-618-55  
Sequence 55, Application US/08485618  
Patent No. 5728533  
GENERAL INFORMATION:  
APPLICANT: Gallatin, W. Michael  
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Seair Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,618  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/173,497  
FILING DATE: 23-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,889  
FILING DATE: 5-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,652  
FILING DATE: 21-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams Jr., Joseph A.  
REGISTRATION NUMBER: 38,659  
REFERENCE/DOCKET NUMBER: 27866/32797  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1161 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein



Db 152 DIAPFLDSSGSSINORDPAQKMDPVKALMGEPASSTLSLMQYNSILKHTFTPEFKNII 211  
Qy 55 DREQIROGLEBELQVLPDGTVMHEGFERASEQIYENRQGYRTA-SVIALTDGELHED 113  
Db 212 DPQSLVDPIVQLQ-----GLTYTATGIRTWMBELFHSKNGSRKSAKILLVITDGOQKRD 266  
Qy 114 LFFYSE--REANSRDGLAIVYCVGVD-FNE-TOLARI-----ADSKDHFPVNDGFOA 164  
Db 267 PLEYSDVIPADKX---GIIRYALGVGDAPQEPALKEMLNTIGSAPQDHFVKVGN-FAA 322  
Qy 165 LOGI 168  
Db 323 LRST 326

RESULT 13  
US-08-482-293A-55  
Sequence 55, Application US/08482293A  
Patent No. 5831029

## GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael  
APPLICANT: Van der Vlieten, Monica  
TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit  
NUMBER OF SEQUENCES: 103  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker Drive, 6300 Sear Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States  
ZIP: 60606-6402

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,293A  
FILING DATE:

## CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652

FILING DATE: 21-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32684

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-482-293A-55

Query Match 15.4%; Score 138.5; DB 2; Length 1161;

Best Local Similarity 28.3%; Pred. No. 6.7e-07;

Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;

Qy 1 DLVFIILKSGSV-LHNNNEIYFVEQLAHKFIISPOLMSFIVSRTGTTMLKTE----- 54

Db 152 DIAPFLDSSGSSINORDPAQKMDPVKALMGEPASSTLSLMQYNSILKHTFTPEFKNII 211

Qy 55 DREQIROGLEBELQVLPDGTVMHEGFERASEQIYENRQGYRTA-SVIALTDGELHED 113  
Db 212 DPQSLVDPIVQLQ-----GLTYTATGIRTWMBELFHSKNGSRKSAKILLVITDGOQKRD 266  
Qy 114 LFFYSE--REANSRDGLAIVYCVGVD-FNE-TOLARI-----ADSKDHFPVNDGFOA 164  
Db 267 PLEYSDVIPADKX---GIIRYALGVGDAPQEPALKEMLNTIGSAPQDHFVKVGN-FAA 322  
Qy 165 LOGI 168  
Db 323 LRST 326

RESULT 14  
US-08-943-363-55  
Sequence 55, Application US/08943363  
Patent No. 5837478

## GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vlieten, Monica

TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

STREET: 233 South Wacker Drive, 6300 Sear Tower

CITY: Chicago

STATE: Illinois

COUNTRY: United States

ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,363

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/173,497

FILING DATE: 23-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,889

FILING DATE: 5-AUG-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652

FILING DATE: 21-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.

REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32684

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-474-6300

TELEFAX: 312-474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-943-363-55

Query Match 15.4%; Score 138.5; DB 2; Length 1161;

Best Local Similarity 28.3%; Pred. No. 6.7e-07;

Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;

Qy 1 DLVFIILKSGSV-LHNNNEIYFVEQLAHKFIISPOLMSFIVSRTGTTMLKTE----- 54

Db 152 DIAPFLDSSGSSINORDPAQKMDPVKALMGEPASSTLSLMQYNSILKHTFTPEFKNII 211

Qy 55 DREQIROGLEBELQVLPDGTVMHEGFERASEQIYENRQGYRTA-SVIALTDGELHED 113

Db 212 DPGSLVDPYVLQ-----GLTYTATGIRTVMEELFHSKNGSRKSAKKILVITTDGQKYRD 266  
QY 114 LFFYSE--REANRSRDIGAIYVCVKD-FNE-TOLARI-----ADSKDHVPVNDGFOA 164  
Db 267 PLEYSDVTPADKA---GIIRYAIGVDAFOEPTALKEINTIGSAPPODHVFKVGN-FAA 322  
QY 165 LOGI 168  
Db 323 LRSI 326

RESULT 15  
US-09-193-043-55  
; Sequence 55: Application US/09193043  
; Patent No. 6251395  
; GENERAL INFORMATION:  
; APPLICANT: Gallatin, Michael W.  
; APPLICANT: Van der Vieren, Monica  
; TITLE OF INVENTION: No. 6251395el Human 2  
; FILE REFERENCE: 27866/35804  
; CURRENT APPLICATION NUMBER: US/09/193, 043  
; EARLIER FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: 08/173, 497  
; EARLIER FILING DATE: 1993-12-23  
; EARLIER APPLICATION NUMBER: 08/286, 889  
; EARLIER FILING DATE: 1994-08-05  
; EARLIER APPLICATION NUMBER: 08/362, 652  
; EARLIER FILING DATE: 1994-12-21  
; EARLIER APPLICATION NUMBER: 08/943, 363  
; EARLIER FILING DATE: 1997-10-03  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 55  
; LENGTH: 1161  
; TYPE: PRT  
; ORGANISM: Rattus rattus  
US-09-193-043-55

Query Match 15.4%; Score 138.5; DB 3; Length 1161;  
Best Local Similarly 28.3%; Pred. No. 6,7e-07;  
Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;  
QY 1 DLYFIIDKSGSV-LHMHNEIYFVEQLAHKEISPOLMSFIVPSTRTIMKLT----- 54  
Db 152 DIAPLIDGSGSINORDPAQMDPFAKALMGEPASTSTLPSLMQYSNLIKTHPTFEKNIL 211  
QY 55 DREQIIRQGLELQKVLPGEDTYMHEGFERASEQIYENROGYRTA-SVIALTDGELHED 113  
Db 212 DPGSLVDPYVLQ-----GLTYTATGIRTVMEELFHSKNGSRKSAKKILVITTDGQKYRD 266  
QY 114 LFFYSE--REANRSRDIGAIYVCVKD-FNE-TOLARI-----ADSKDHVPVNDGFOA 164  
Db 267 PLEYSDVTPADKA---GIIRYAIGVDAFOEPTALKEINTIGSAPPODHVFKVGN-FAA 322  
QY 165 LOGI 168  
Db 323 LRSI 326

Search completed: June 13, 2005, 20:05:09  
Job time: 16.9024 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 20:03:29 / Search time 55.4703 Seconds  
(without alignments)  
1195.537 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_44\_216  
Perfect score: 899  
Sequence: 1 DLVFILDKSGSVLHNMNEY.....HFPVNDGQALQGIHSIL 173

Scoring table: BLOSUM62  
Gapco 10.0, Gapext 0.5

Searched: 1710399 seqs, 383334425 residues  
Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubppa/PTCTUS\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubppa/PTCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubppa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubppa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubppa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubppa/US10C\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubppa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubppa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	899	100.0	328	US-10-038-307-26 Sequence 26, App1
2	899	100.0	328	US-10-201-292-26 Sequence 26, App1
3	899	100.0	333	US-09-796-753-12 Sequence 12, App1
4	899	100.0	333	US-10-038-307-2 Sequence 2, App1
5	899	100.0	333	US-10-201-292-2 Sequence 22, App1
6	899	100.0	342	US-10-038-307-22 Sequence 22, App1
7	899	100.0	342	US-10-201-292-22 Sequence 22, App1
8	899	100.0	345	US-10-038-307-24 Sequence 24, App1
9	899	100.0	345	US-10-201-292-24 Sequence 24, App1
10	899	100.0	403	US-09-833-245-621 Sequence 621, App1
11	899	100.0	460	US-10-201-292-28 Sequence 28, App1

12	899	100.0	460	US-10-201-292-30 Sequence 30, App1
13	899	100.0	479	US-10-201-292-32 Sequence 32, App1
14	899	100.0	504	US-10-201-292-34 Sequence 34, App1
15	899	100.0	529	US-10-201-292-36 Sequence 36, App1
16	899	100.0	538	US-10-047-542-99 Sequence 99, App1
17	899	100.0	551	US-10-038-307-18 Sequence 18, App1
18	899	100.0	551	US-10-201-292-18 Sequence 18, App1
19	899	100.0	564	US-09-918-715-187 Sequence 187, App
20	899	100.0	564	US-09-918-715-232 Sequence 232, App
21	899	100.0	564	US-10-038-307-20 Sequence 20, App1
22	899	100.0	564	US-10-201-292-20 Sequence 20, App1
23	899	100.0	564	US-10-301-822-199 Sequence 199, App
24	899	100.0	564	US-10-408-765A-1823 Sequence 1823, App
25	899	100.0	564	US-10-474-794-187 Sequence 187, App
26	899	100.0	564	US-10-474-794-232 Sequence 232, App
27	899	99.4	403	US-09-833-245-620 Sequence 620, App
28	899	99.0	562	US-09-918-715-194 Sequence 194, App
29	899	99.0	562	US-09-918-715-301 Sequence 301, App
30	899	99.0	562	US-10-474-794-301 Sequence 301, App
31	899	99.0	562	US-10-474-794-301 Sequence 301, App
32	899	97.7	534	US-10-038-307-12 Sequence 12, App1
33	899	97.7	534	US-10-038-307-12 Sequence 12, App1
34	899	97.7	543	US-10-038-307-10 Sequence 10, App1
35	899	97.7	543	US-10-038-307-14 Sequence 14, App1
36	899	97.7	543	US-10-038-307-16 Sequence 16, App1
37	899	97.7	543	US-10-201-292-14 Sequence 14, App1
38	899	97.7	543	US-10-201-292-14 Sequence 14, App1
39	899	97.7	543	US-10-201-292-16 Sequence 16, App1
40	899	57.7	488	US-10-038-307-6 Sequence 6, App1
41	899	57.7	488	US-10-201-292-6 Sequence 6, App1
42	899	57.7	488	US-10-201-292-6 Sequence 6, App1
43	899	57.7	488	US-10-368-087-16 Sequence 16, App1
44	899	57.7	488	US-10-104-047-2639 Sequence 2639, App
45	899	57.5	487	US-09-796-753-54 Sequence 54, App1

ALIGNMENTS

RESULT 1				
US-10-038-307-26				
Sequence 26, Application US/10038307				
Publication No. US20030134786A1				
GENERAL INFORMATION:				
APPLICANT: James B. ROTTMAN				
APPLICANT: Theresa L. O'KEEFE				
APPLICANT: Engin OZKANNAK				
APPLICANT: Judith V. HEALEY				
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods				
FILE REFERENCE: 7853-253-999				
CURRENT APPLICATION NUMBER: US/10/038,307				
CURRENT FILING DATE: 2002-06-28				
NUMBER OF SEQ ID NOS: 26				
SOFTWARE: FASTSEQ for Windows Version 4.0				
SEQ ID NO 26				
LENGTH: 328				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-10-038-307-26				
Query Match				
Best Local Similarity 100.0%; Score 899; DB 14; Length 328;				
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	DLVFILDKSGSVLHNMNEYIFVPEQLAKRISPOLRMSFIYFSTRGTTIMKLTEDRQIR	60	
DB	44	DLVFILDKSGSVLHNMNEYIFVPEQLAKRISPOLRMSFIYFSTRGTTIMKLTEDRQIR	103	
QY	61	OGAEELQKVPFGGDTYHGEGERASEQIYYENRGVYTAATIALTDGELHEDLFFYSER	120	
DB	104	OGAEELQKVPFGGDTYHGEGERASEQIYYENRGVYTAATIALTDGELHEDLFFYSER	163	
QY	121	ENANRSDLAIVYCVGVKDNENETOLARIADSKOHVFPVNDGQALQGIHSIL	173	

```
Db      164 EARNRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 216

RESULT 2
US-10-201-292-26
; Sequence 26, Application US/10201292
; Publication No. US2003014193A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/201,292
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-201-292-26

Query Match      100.0%; Score 899; DB 14; Length 328;
Best Local Similarity 100.0%; Pred. No. 1e-86;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DLYFIIDKSGSVLHNMNIYYFVEQLAHKFIISPOLRMSFVIFSTRGTTLMKLTEDRQIR 60
        |||||||
Db      44 DLYFIIDKSGSVLHNMNIYYFVEQLAHKFIISPOLRMSFVIFSTRGTTLMKLTEDRQIR 103
        |||||||

Qy      61 QGLEELQKVLPGGDTYMHGFEFASQIYYENRGYRTASVIALTDGELHEDLFFYSER 120
        |||||||
Db      104 QGLEELQKVLPGGDTYMHGFEFASQIYYENRGYRTASVIALTDGELHEDLFFYSER 163
        |||||||

Qy      121 EARNRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 173
        |||||||
Db      164 EARNRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 216
        |||||||

RESULT 3
US-09-796-753-12
; Sequence 12, Application US/09796753
; Publication No. US2003027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
```

```
        ; PRIOR FILING DATE: 1999-07-30
        ; PRIOR APPLICATION NUMBER: 09/399,723
        ; PRIOR FILING DATE: 1999-09-20
        ; PRIOR APPLICATION NUMBER: 09/409,634
        ; PRIOR FILING DATE: 1999-09-30
        ; PRIOR APPLICATION NUMBER: 09/471,179
        ; PRIOR FILING DATE: 1999-12-23
        ; PRIOR APPLICATION NUMBER: 09/474,071
        ; PRIOR FILING DATE: 1999-12-29
        ; PRIOR APPLICATION NUMBER: 09/474,072
        ; PRIOR FILING DATE: 1999-12-29
        ; PRIOR APPLICATION NUMBER: 09/514,010
        ; PRIOR FILING DATE: 2000-02-25
        ; PRIOR APPLICATION NUMBER: 09/516,745
        ; PRIOR FILING DATE: 2000-03-01
        ; PRIOR APPLICATION NUMBER: 09/572,002
        ; PRIOR FILING DATE: 2000-05-14
        ; PRIOR APPLICATION NUMBER: 09/597,993
        ; PRIOR FILING DATE: 2000-06-19
        ; PRIOR APPLICATION NUMBER: 09/599,596
        ; PRIOR FILING DATE: 2000-06-22
        ; PRIOR APPLICATION NUMBER: 09/630,334
        ; PRIOR FILING DATE: 2000-07-31
        ; PRIOR APPLICATION NUMBER: 09/606,565
        ; PRIOR FILING DATE: 2000-06-29
        ; PRIOR APPLICATION NUMBER: 09/606,317
        ; PRIOR FILING DATE: 2000-06-29
        ; PRIOR APPLICATION NUMBER: 09/665,666
        ; PRIOR FILING DATE: 2000-09-20
        ; PRIOR APPLICATION NUMBER: 09/677,751
        ; PRIOR FILING DATE: 2000-09-30
        ; NUMBER OF SEQ ID NOS: 162
        ; SEQ ID NO 12
        ; LENGTH: 333
        ; TYPE: PRT
        ; ORGANISM: Homo sapiens
US-09-796-753-12

Query Match      100.0%; Score 899; DB 10; Length 333;
Best Local Similarity 100.0%; Pred. No. 1e-86;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DLYFIIDKSGSVLHNMNIYYFVEQLAHKFIISPOLRMSFVIFSTRGTTLMKLTEDRQIR 60
        |||||||
Db      44 DLYFIIDKSGSVLHNMNIYYFVEQLAHKFIISPOLRMSFVIFSTRGTTLMKLTEDRQIR 103
        |||||||

Qy      61 QGLEELQKVLPGGDTYMHGFEFASQIYYENRGYRTASVIALTDGELHEDLFFYSER 120
        |||||||
Db      104 QGLEELQKVLPGGDTYMHGFEFASQIYYENRGYRTASVIALTDGELHEDLFFYSER 163
        |||||||

Qy      121 EARNRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 173
        |||||||
Db      164 EARNRSDLGAIYCVGVKDFNETQLARIADSKDHVFPVNDGFQALQGIHSIL 216
        |||||||

RESULT 4
US-10-038-307-2
; Sequence 2, Application US/10038307
; Publication No. US20030134786A1
; GENERAL INFORMATION:
; APPLICANT: James B. ROTTMAN
; APPLICANT: Theresa L. O'KEEFE
; APPLICANT: Engin OZKAYNAK
; APPLICANT: Judith J. HEALEY
; TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods
; FILE REFERENCE: 7853-253-999
; CURRENT APPLICATION NUMBER: US/10/038,307
; CURRENT FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 333
; TYPE: PRT
```

ORGANISM: Homo sapiens  
US-10-038-307-2

Query Match 100.0%; Score 899; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 216

## RESULT 5

US-10-201-292-2  
Sequence 2, Application US/10201292  
Publication No. US20030144193A1

GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201,292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-2

Query Match 100.0%; Score 899; DB 14; Length 333;  
Best Local Similarity 100.0%; Pred. No. 1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 216

## RESULT 6

US-10-038-307-22  
Sequence 22, Application US/10038307  
Publication No. US20030134786A1

GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038,307  
CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 22  
LENGTH: 342  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-307-22

Query Match 100.0%; Score 899; DB 14; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 53 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 112  
QY 61 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 113 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 172  
QY 121 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 173  
DB 173 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 225

## RESULT 7

US-10-201-292-22  
Sequence 22, Application US/10201292  
Publication No. US20030144193A1

GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201,292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 342  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-22

Query Match 100.0%; Score 899; DB 14; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 53 DLYFIIDKSGSVLHHNNEIYFVQQLAKHFIISPOLRMSFIVFSTRTGTTLMKLTEDRQIR 112  
QY 61 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 113 QGLEELQKVLPGGDTYHHEGFERASBOIYYENRQGYRTASVIALTDGELHEDLFFYSER 172  
QY 121 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 173  
DB 173 EANRSRDLAGIIVCVGVKDFNETOLARIADSKOHVFPVNDGFOALQGIHSIL 225

## RESULT 8

US-10-038-307-24  
Sequence 24, Application US/10038307  
Publication No. US20030134786A1

GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/038,307

CURRENT FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 345  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-10-038-307-24

Query Match 100.0%; Score 899; DB 14; Length 345;  
Best Local Similarity 100.0%; Pred. No. 1.1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 216

## RESULT 9

US-10-201-292-24  
Sequence 24, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT FILING DATE: 2003-02-14  
CURRENT APPLICATION NUMBER: US/10/201,292  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 345  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-10-201-292-24

Query Match 100.0%; Score 899; DB 14; Length 345;  
Best Local Similarity 100.0%; Pred. No. 1.1e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 216

## RESULT 10

US-09-833-245-621  
Sequence 621, Application US/09833245  
Publication No. US20040010134A1  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Albumin Fusion Proteins  
FILE REFERENCE: PFS46PCF  
CURRENT APPLICATION NUMBER: US/09/833,245

CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: 60/229, 358  
PRIOR FILING DATE: 2000-04-12  
PRIOR APPLICATION NUMBER: 60/256, 931  
PRIOR FILING DATE: 2000-12-21  
PRIOR APPLICATION NUMBER: 60/199, 384  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 2267  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 621  
LENGTH: 403  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-09-833-245-621

Query Match 100.0%; Score 899; DB 11; Length 403;  
Best Local Similarity 100.0%; Pred. No. 1.4e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 216

## RESULT 11

US-10-201-292-28  
Sequence 28, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT FILING DATE: 2003-02-14  
CURRENT APPLICATION NUMBER: US/10/201,292  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 460  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-10-201-292-28

Query Match 100.0%; Score 899; DB 14; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.6e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 60  
DB 44 DLYFIIDKSGSVLHNNHNIYFVEQLAHKFIISPOLMSFIVFSTRTGTTLMKLTEDRQIR 103  
QY 61 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 QGLEELQKVLPGGDTYMHGFEFASBQIYYENRQGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 173  
DB 164 EANRSRDLAGAIYVCVGKDFNETQLARIADSKDHVPFVNDGFOALQGIHSIL 216

## RESULT 12

US-10-201-292-30  
Sequence 30, Application US/10201292

Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 30  
LENGTH: 460  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-30

Query Match 100.0%; Score 899; DB 14; Length 460;  
Best Local Similarity 100.0%; Pred. No. 1.6e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 60  
DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 103  
QY 61 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 173  
DB 164 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 216

RESULT 13  
US-10-201-292-32  
Sequence 32, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 32  
LENGTH: 479  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-32

Query Match 100.0%; Score 899; DB 14; Length 479;  
Best Local Similarity 100.0%; Pred. No. 1.7e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 60  
DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 103  
QY 61 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 173  
DB 164 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 216

RESULT 14  
US-10-201-292-34  
Sequence 34, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34  
LENGTH: 504  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-34

Query Match 100.0%; Score 899; DB 14; Length 504;  
Best Local Similarity 100.0%; Pred. No. 1.8e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 60  
DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 103  
QY 61 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 173  
DB 164 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 216

RESULT 15  
US-10-201-292-36  
Sequence 36, Application US/10201292  
Publication No. US20030144193A1  
GENERAL INFORMATION:  
APPLICANT: James B. ROTTMAN  
APPLICANT: Theresa L. O'KEEFE  
APPLICANT: Engin OZKAYNAK  
APPLICANT: Judith J. HEALEY  
TITLE OF INVENTION: Tango 197 and Tango 216 Compositions and Methods  
FILE REFERENCE: 7853-253-999  
CURRENT APPLICATION NUMBER: US/10/201.292  
CURRENT FILING DATE: 2003-02-14  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 36  
LENGTH: 529  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-201-292-36

Query Match 100.0%; Score 899; DB 14; Length 529;  
Best Local Similarity 100.0%; Pred. No. 2e-86;  
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 60  
DB 44 DLYFILDKSGSVLHNMNEIYFVEQLAHKFI SPQLRMSFIVFSTRGTTMLKLTEDRQIR 103  
QY 61 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 120  
DB 104 OGLEELQKVLPGGDTYHHEGFERASEQIYYENRGYRTASVIALTDGELHEDLFFYSER 163  
QY 121 EANRSRDGAIVYCVGVKDFNETOLARIADSKDHVFPVNDGFOALQGIHSIL 173

Db 164 EARNRSDLGAIVYCVGVKDFNFTQLARIADSKOHVPVNDGFQALOGIHSIL 216

Search completed: June 13, 2005, 20:36:51  
Job time : 56.4703 secs

---



B.  
 A:Reference number: A31108; MUID:88315033; PMID:2457584  
 A:Accession: A31108  
 A:Molecule type: mRNA  
 A:Residues: 1-1153 <COR>  
 A:Note: references: UNIPROT:P11215; GB:J03925; NID:G187284; PIDN:AAA59544.1; PID:G307148  
 A:Note: part of this sequence was confirmed by protein sequencing  
 R:Arnaut, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.  
 J. Cell Biol. 106, 2153-2158, 1988  
 A:Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M  
 A:Reference number: A28915; MUID:88257215; PMID:2454931  
 A:Accession: A28915  
 A:Molecule type: mRNA  
 A:Residues: 1-499,501-965, 'P', 967-1153 <ARN>  
 A:Cross-references: GB:M18044; GB:U03270; GB:M19664; GB:X07421; NID:G186935; PIDN:AAA594  
 A:Note: the authors translated the codon TAC for residue 1129 as Thr  
 A:Note: part of this sequence, including the amino end of the mature protein, was confir  
 R:Shelley, C.S.; Arnaut, M.A.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 10525-10529, 1988  
 A:Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg  
 A:Reference number: A41600; MUID:92073318; PMID:1683702  
 A:Accession: A41600  
 A:Molecule type: DNA  
 A:Residues: 1-9 <SHE>  
 A:Cross-references: GB:M76724; NID:G180018; PIDN:AAA58410.1; PID:G553215  
 R:Arnaut, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988  
 A:Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion  
 A:Reference number: A94193; MUID:88190151; PMID:2833753  
 A:Accession: A30892  
 A:Molecule type: mRNA  
 A:Residues: 917-1042 <AR2>  
 A:Cross-references: GB:M18044  
 R:Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.W.; Back, A.L.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989  
 A:Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor  
 A:Reference number: A32218; MUID:89098893; PMID:2553162  
 A:Accession: A32218  
 A:Molecule type: mRNA  
 A:Residues: 9-1153 <HIC>  
 A:Cross-references: GB:U04145; NID:G189068; PIDN:AAA59903.1; PID:G386795  
 A:Note: the last three bases of intron 13, CAG, are included in some but not all mature  
 R:Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.  
 J. Immunol. 150, 480-490, 1993  
 A:Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-  
 A:Reference number: A46526; MUID:93123748; PMID:8419480  
 A:Accession: A46526  
 A:Molecule type: DNA  
 A:Status: not compared with conceptual translation  
 A:Residues: 1-499,501-1153 <FLB>  
 A:Cross-references: GB:S52227; NID:G263047; PIDN:AAA24821.1; PID:G263049  
 A:Note: the last three bases of intron 13, CAG, are included in some but not all mature  
 R:Note: sequence extracted from NCBI backbone (NCBI:P:121963)  
 R:Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaut, M.A.  
 Biochim. Biophys. Acta 874, 368-371, 1986  
 A:Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across sp  
 A:Reference number: A90664; MUID:87076671; PMID:35359202  
 A:Accession: A26091  
 A:Molecule type: protein  
 A:Residues: 17-31 <PIE>  
 A:Experimental source: granulocytes  
 R:Pahl, H.L.; Rossmartin, A.G.; Tenen, D.G.  
 Blood 79, 865-870, 1992  
 A:Title: Characterization of the myeloid-specific CD11b promoter.  
 A:Reference number: 152567; MUID:92144986; PMID:1346576  
 A:Accession: 152567  
 A:Status: translated from GB/EMBL/DBD  
 A:Molecule type: DNA  
 A:Residues: 1-9 <RES>  
 A:Cross-references: GB:M84477; NID:G180184; PIDN:AAA51960.1; PID:G553219  
 C:Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1  
 C:Genetic:

A:Gene: GDB:ITGM, CR3A  
 A:Cross-references: GDB:120599; OMIM:120980  
 A:Map position: 16p11.2-16p11.2  
 A:Note: promoter contains a GATA motif and two Sp1 consensus binding sites  
 C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
 C:Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; magr  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>  
 F:17-1153/Domain: extracellular #status predicted <EXT>  
 F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F:465-473/Region: calcium/magnesium binding #status predicted  
 F:530-538/Region: calcium/magnesium binding #status predicted  
 F:593-601/Region: calcium/magnesium binding #status predicted  
 F:1109-1134/Domain: transmembrane #status predicted <TM>  
 F:1135-1153/Domain: intracellular #status predicted <INT>  
 F:86,240,391,469,693,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding  
 Query Match 14.3% Score 128.5; DB 1; Length 1153;  
 Best Local Similarity 26.7%; Pred. No. 0.0042;  
 Matches 51; Conservative 38; Mismatches 63; Indels 39; Gaps 11;  
 QY 1 DLYFIIDKSGSVL-HHMEIYFVEQLAHKFIISPOLMSFIVSTGTTMLKTED---- 55  
 Db 150 DIAFLIDGSGSIIIPHFRMKFVST-----VMEQLKSKTLFS-----LMQYSEERRH 199  
 QY 56 -----RQIQGLLELOKVLPGDITVHSEPERASEDIYENRQVFTA-SVITALT 106  
 Db 200 FTFKERNNNPNRSLVKVPIYOL--GRITVATGIRKVRRELFTITGARKNAKFLIVIT 257  
 QY 107 DGLIHEDLFPYSE--REANSRDLGAIIVCVGKDNENQTLAR-----INDS--KDHVFP 157  
 Db 258 DGEKFDPLGYEVITREADRE--GVIRYIVGDAFRSEKSHQELNTTIAKSPRDHVFQ 314  
 QY 158 VNDGFQALQGI 168  
 Db 315 VNN-PEALKTI 324  
 RESULT 3  
 A40020  
 collagen alpha 1(XII) chain precursor - chicken  
 N:Alternate names: fibrochimerin  
 C:Species: Gallus gallus (chicken)  
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #ext change 09-Jul-2004  
 C:Accession: A40020; A34485; B34485; A28037; S23814; S22254; S28811  
 R:Yamagata, M.; Yamada, K.M.; Yamada, S.S.; Shinomura, T.; Tanaka, H.; Nishida, Y.; Obari  
 J. Cell Biol. 115, 209-221, 1991  
 A:Title: The complete primary structure of type XII collagen shows a chimeric molecule w  
 A:Reference number: A40020; MUID:92011862; PMID:1918137  
 A:Accession: A40020  
 A:Molecule type: mRNA  
 A:Residues: 1-3124 <YAM>  
 A:Cross-references: UNIPROT:P13944; GB:D00824; NID:G222810; PIDN:BA400701.1; PID:G222811  
 A:Note: in the authors' translation residues 1216-1219 are shown after residue 1235 and,  
 R:Gordon, M.K.; Gerecke, D.R.; Duddle, B.; van der Reek, M.; Olsen, B.R.  
 J. Biol. Chem. 264, 19772-19778, 1989  
 A:Title: Type XII collagen. A large multidomain molecule with partial homology to type I  
 A:Reference number: A34485; MUID:90062079; PMID:2584192  
 A:Accession: A34485  
 A:Molecule type: mRNA  
 A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>  
 A:Cross-references: EMBL:J05137; NID:G211284; PIDN:AAA48635.1; PID:G211285  
 A:Accession: B34485  
 A:Molecule type: protein  
 A:Residues: 2772-2792; 2846-2873 <GOR2>  
 R:Gordon, M.K.; Gerecke, D.R.; Olsen, B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987  
 A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c  
 A:Reference number: A28037; MUID:87317590; PMID:3476925  
 A:Accession: A28037  
 A:Molecule type: mRNA  
 A:Residues: 2960-2976, 'F', 2978-3074, 'AG' <GOR3>



A/Cross-references: EMBL:M17375; NID:G211649; PIDN:AAA8718.1; PID:G211650  
A/Note: this sequence has been revised in reference A34485  
R/Koch, M.; Bernasconi, C.; Chiquet, M.  
Eur. J. Biochem. 207, 847-856, 1992  
A/Title: A major oligomeric fibronectin proteoglycan identified as a novel large form of  
A/Reference number: S28811; MUID:92362621; PMID:13323460  
A/Accession: S28811  
A/Molecule type: protein  
A/Residues: 'X', 1333, 'Q', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517  
R/Dublet, B.; van der Reek, M.  
J. Biol. Chem. 267, 17724-17727, 1992  
A/Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of peptide  
A/Reference number: S22254; MUID:88087065; PMID:3121603  
A/Accession: S22254  
A/Molecule type: protein  
A/Residues: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DUB>  
R/Trieb, J.; Trieb, B.  
Biochim. Biophys. Acta 1171, 97-98, 1992  
A/Title: The two splice variants of collagen XII share a common 5' end.  
A/Reference number: S28811; MUID:93042014; PMID:1420368  
A/Accession: S28811  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>  
A/Cross-references: EMBL:X67327  
C/Genetics:  
A/Insertions: 2845/3; 2863/3; 2899/3; 2922/1; 2985/1; 3008/1; 3065/1  
C/Keywords: alternative splicing; cell binding; coll; connective tissue; disulfide  
P/1-23/Domain: signal sequence #status predicted <SIG>  
P/24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>  
P/24, 1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted  
P/24-114/Domain: IIR #status predicted <IIR>  
P/24-105/Domain: fibronectin type III repeat homology <FN3A>  
P/137-301/Domain: von Willebrand factor type A repeat homology <VWA1>  
P/332-425/Domain: IIR #status predicted <IIR>  
P/332-614/Domain: fibronectin type III repeat homology <FN3B>  
P/437-601/Domain: von Willebrand factor type A repeat homology <VWA2>  
P/629-1178/Domain: IIR #status predicted <IIR>  
P/630-711/Domain: fibronectin type III repeat homology <FN3C>  
P/721-802/Domain: fibronectin type III repeat homology <FN3D>  
P/812-895/Domain: fibronectin type III repeat homology <FN3E>  
P/905-986/Domain: fibronectin type III repeat homology <FN3F>  
P/995-1076/Domain: fibronectin type III repeat homology <FN3G>  
P/1086-1169/Domain: fibronectin type III repeat homology <FN3H>  
P/1197-1361/Domain: von Willebrand factor type A repeat homology <VWA3>  
P/1384-2296/Domain: IIR #status predicted <IIR>  
P/1384-1466/Domain: fibronectin type III repeat homology <FN3I>  
P/1474-1557/Domain: fibronectin type III repeat homology <FN3J>  
P/1566-1647/Domain: fibronectin type III repeat homology <FN3K>  
P/1655-1738/Domain: fibronectin type III repeat homology <FN3L>  
P/1756-1838/Domain: fibronectin type III repeat homology <FN3M>  
P/1847-1928/Domain: fibronectin type III repeat homology <FN3N>  
P/1937-2019/Domain: fibronectin type III repeat homology <FN3O>  
P/2028-2110/Domain: fibronectin type III repeat homology <FN3P>  
P/2119-2199/Domain: fibronectin type III repeat homology <FN3Q>  
P/2207-2294/Domain: fibronectin type III repeat homology <FN3R>  
P/2325-2499/Domain: von Willebrand factor type A repeat homology <VWA4>  
P/2438-2440/Domain: cell adhesion #status predicted  
P/2509-2750/Domain: IIR, homologous to NC2 domain of type IX collagen #status predicted  
P/2751-2902/Domain: cell attachment (R-G-D) motif  
P/2899-2901/Domain: collagenous COL2 #status predicted <COL2>  
P/2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>  
P/2946-3048/Domain: collagenous COL1 #status predicted <COL1>  
P/3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>  
P/321, 1006, 1032, 1044, 1512, 1767, 2210, 2273, 2532, 2683/Binding site: carbohydrate (Asn) (cova  
P/2780, 2789, 2832, 2842, 2860, 2866, 2869, 3004, 3007/Modified site: hydroxyproline (Pro) #stat

Db 2327 DLYFLDKSGSV-LHMHNEIYFVEQLAHKF-ISPQ-LRMSFVFGTGTMLKLT--E 54  
Qy 55 DREQIRQGLSELQVLPDGTVMHEGPERASEQIY-YENRGYRTASVIALTDGELHED 113  
Db 2387 DKQALALQNVQ--YHGNTTRTKALFLPEKYLTMESGRRBVPKVLVVTGGRSQDE 2444  
Qy 114 LFFYSERERANSRDLGAIYVCVGVKDNFETQLARIAD--SKDHFPPVNDGQALQGI 168  
Db 2445 V---RRAAVTVIQSGSGSVFVGVADVADYNELAKIASKPSERHFIYDD-FDAFEKI 2496  
RESULT 4  
A45638  
1Immunodominant microneme protein Etp100 - Eimeria tenella  
C/Species: Eimeria tenella  
C/Date: 22-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C/Accession: A45638  
R/Tomley, F.M.; Clarke, L.B.; Kawazoe, U.; Djikema, R.; Kok, J.J.  
Mol. Biochem. Parasitol. 49, 277-288, 1991  
A/Title: Sequence of the gene encoding an immunodominant microneme protein of Eimeria te.  
A/Reference number: A45638; MUID:92131064; PMID:1775171  
A/Accession: A45638  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-712 <TOM>  
A/Cross-references: UNIPROT:O43981; GB:A033905; GB:M73495; NID:G2707732; PIDN:AAD03350.  
A/Note: sequence extracted from NCBI backbone (NCBI:7752, NCBI:7756)  
P/48-218/Domain: von Willebrand factor type A repeat homology <VWA1>  
P/238-286/Domain: thrombospondin type 1 repeat homology <THR1>  
P/309-371/Domain: thrombospondin type 1 repeat homology <THR2>  
P/372-432/Domain: thrombospondin type 1 repeat homology <THR3>  
P/433-493/Domain: thrombospondin type 1 repeat homology <THR4>  
P/494-556/Domain: thrombospondin type 1 repeat homology <THR5>  
P/560-610/Domain: thrombospondin type 1 repeat homology <THR6>  
Query Match 13.5%; Score 121.5; DB 2; Length 712;  
Best Local Similarity 25.6%; Pred. No. 0.0097;  
Matches 43; Conservative 33; Mismatches 59; Indels 33; Gaps 9;  
Qy 1 DLYFLDKSGSV-LHMHNEIYFVEQLAHKF-ISPQ-LRMSFVFGTGTMLKLTEDRE 57  
Db 50 DVLMLVBSGSGTGRNKRQFLIEDVNSMPSIDVDRGLTFAI-----RS 98  
Qy 58 QIRQGLSELQVLPDGTVMHEGPERASEQIYENRGYR--TASVI 102  
Db 99 KVRNMLSDPRATNSLAIASARSLSTGVLYTHYGLQDA-KKLLYTNMGARRNVPCLV 157  
Qy 103 IALTDELHEDLFFYSERERANSRDLGAIYVCVGV-KDFNETQLARIA 149  
Db 158 LVMTDGA--SVLPQTRSSAALRDAGAIIVLVGVSGVNSSECRSIA 203  
RESULT 5  
RWHUIC  
cell surface glycoprotein CD11c precursor - human  
C/Alternate names: leukocyte adhesion receptor p150,95 alpha chain  
C/Species: Homo sapiens (man)  
C/Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 09-Jul-2004  
C/Accession: A35543; A35543; S00864  
R/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 12750-12751, 1990  
A/Reference number: A35584  
A/Contents: erratum  
A/Accession: A35584  
A/Molecule type: DNA  
A/Residues: 1-1163 <COR>  
A/Cross-references: UNIPROT:P20702  
A/Note: this revision to the sequence from reference A35543 includes the carboxyl end  
R/Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.  
J. Biol. Chem. 265, 2782-2788, 1990  
A/Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.  
A/Reference number: A35543; MUID:90153906; PMID:2303426  
A/Accession: A35543

A/Molecule type: DNA  
 A/Residues: 1-834 <CO2>  
 A/Note: this sequence has been revised in reference A3584  
 R/Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.  
 EMBL J. 6, 4023-4028, 1987  
 A/Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte  
 A/Reference number: S00864; MUID:88166645; PMID:3327687  
 A/Accession: S00864  
 A/Molecule type: mRNA  
 A/Residues: 1-755; 'L', 757-1163 <CO3>  
 A/Cross-references: GB:M1695; EMBL:Y00093; NID:G487829; PIND:AAA59180.1; PID:G487830  
 A/Note: part of this sequence was confirmed by protein sequencing  
 C/Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my  
 C/Genetics:  
 A/Gene: GDB:ITGAX; CD11C  
 A/Cross-references: GDB:119758; OMIM:151510  
 A/Map position: 16p11.2-16p11.2  
 C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat hom  
 C/Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>  
 F:20-1107/Domain: extracellular #status predicted <EXT>  
 F:149-319/Domain: von Willebrand factor type A repeat homology <VWA4>  
 F:1108-1133/Domain: transmembrane #status predicted <TM>  
 F:1134-1163/Domain: intracellular #status predicted <INT>  
 F:61,89,392,697,735,899,939,1050/Binding site: carboxylate (Asn) (covalent) #status pre

Query Match 13.5%; Score 121.5; DB 1; Length 1163;  
 Best Local Similarity 25.1%; Pred. No. 0.017;  
 Matches 47; Conservative 35; Mismatches 74; Indels 31; Gaps 9;

QY 1 DIVYILDKSGSV-LAHNNEIYFVEQLAHKISPOLMSFVSTRTGTLTKLDEDE-- 57  
 D 151 DIVYILDKSGSVLSRNPATMNFRAVISOQRFSTPSLMQFSNKFQTHFEFFRTS 210  
 QY 58 ---QIRGQLEBLQVLPGSDTYMEGPERASEQIYENRGQYRAS-VIATLTDEGLHED 113  
 D 211 NPLSLASVHQLQ-----GFTYTAATQNVNHRFLPHASYGARRATKILVITDGKEGD 265  
 QY 114 LFYSESRANRSDLAIVYCVGK-----KDFNETQLARIAD-SKHVFPVNDG 161  
 D 266 SLDYKD-VIPMADAAGIIRYAVIGLAFQNRNSWKEIND-----LASKPSQEHIFKVED- 318  
 QY 162 FOALQGI 168  
 D 319 FDALKDI 325

## RESULT 6

S00551

leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse  
 N/Alternate names: complement-3 receptor alpha chain  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
 C/Accession: S00551; 159078

R/Pytela, R.  
 EMBL J. 7, 1371-1378, 1988  
 A/Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the  
 A/Reference number: S00551; MUID:88312584; PMID:3044779  
 A/Accession: S00551  
 A/Molecule type: DNA  
 A/Residues: 1-1153 <PYT>

A/Cross-references: UNIPROT:P05555; EMBL:X07640; NID:G52982; PIND:CAA30479.1; PID:G52983  
 A/Note: the authors translated the codon CAC for residue 569 as Gln  
 R/Saetre, L.; Roman, J.M.; Teplov, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,  
 Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986  
 A/Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recep  
 A/Reference number: 159078; MUID:86287312; PMID:2942940  
 A/Accession: 159078  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 11-44 <RES>

A/Cross-references: GB:M14293; NID:g198993; PIND:AAA3484.1; PID:g554193

C/Genetics:  
 A/Gene: Mac-1  
 C/Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo  
 C/Keywords: cell adhesion; glycoprotein; transmembrane protein  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental  
 F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F:1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 13.4%; Score 120.5; DB 2; Length 1153;  
 Best Local Similarity 27.2%; Pred. No. 0.021;  
 Matches 52; Conservative 34; Mismatches 66; Indels 39; Gaps 12;

QY 1 DIVYILDKSGSVLAHNNIYFVEQLAHKIS---POLMSFVSTRTGTLTKLDEDE-- 55  
 D 150 DIVYILDKSGSVI-----NNIDF---QKKKEFVSTVMEQFKSKTFLS-----LMQYSDFR 197  
 QY 56 ---REQIRGQLEBLQVLPGSDTYMEGPERASEQIYENRGQYRAS-VIATLTDEGLHED 106  
 D 198 IHTFTNDFKKNPSPRSVSIKQLNGRTKTASGIRKVRKELFKTNGARENAAKILVIT 257  
 QY 107 DGEIHEDLFPYSE-REANRSDLAIVYCVGKVD-FNETQLARIAD-----SKHVFP 157  
 D 258 DGEKFGDPLDYKQVPEADRA---GVIRYVIGVGNAPNKFQSRRELDITASKPAGHVFQ 314  
 QY 158 VNDGFOALQGI 168  
 D 315 V-DNFEALNTI 324

## RESULT 7

C2MS

classical-complement-pathway C3/C5 convertase (EC 3.4.21.43) C2 component precursor - mor  
 N/Alternate names: C3 convertase; C5 convertase; complement C2  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
 C/Accession: A38876; B36593; I54429

R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
 submitted to Genbank, January 1991  
 A/Reference number: A38875  
 A/Accession: A38876  
 A/Molecule type: DNA

A/Residues: 1-760 <IS2>  
 A/Cross-references: UNIPROT:P21180; GB:M57891; GB:J05661; NID:g192436; PIND:AAA63294.1; I  
 R/Ishikawa, N.; Nonaka, M.; Wetzel, R.A.; Colten, H.R.  
 J. Biol. Chem. 265, 19040-19046, 1990  
 A/Title: Murine complement C2 and factor B genomic and cDNA cloning reveals different me

A/Reference number: A36593; MUID:91035430; PMID:2229060  
 A/Accession: B36593  
 A/Molecule type: mRNA

A/Residues: 1-760 <ISH>  
 A/Cross-references: EMBL:M57891; NID:g192436; PIND:AAA63294.1; PID:g192437  
 R/Falut, A.; Wakeland, E.K.; McConnell, T.J.; Gitlin, J.; Whitehead, A.S.; Colten, H.R.  
 Immunogenetics 25, 290-298, 1987  
 A/Title: DNA polymorphism of MHC III genes in inbred and wild mouse strains.  
 A/Reference number: I54429; MUID:87192938; PMID:2883115  
 A/Accession: I54429

A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Residues: 660-677; 'R', 679,681-723; 'G', 725 <RES>  
 A/Molecule type: mRNA  
 A/Residues: 660-677; 'R', 679,681-723; 'G', 725 <RES>  
 A/Cross-references: GB:M16271; NID:g199289; PIND:AAA3562.1; PID:g199290  
 C/Genetics:

A/Intons: 16/1; 91/1; 153/1; 212/1; 245/3; 290/3; 337/1; 384/1; 414/1; 461/1; 492/3; 53;  
 C/Complex: The proenzyme forms a complex with C4a and is activated by cleavage into C2a,  
 C/Function:  
 A/Description: cleaves complement C3 and complement C5 alpha chains  
 C/Superfamily: complement B/C2; complement factor H repeat homology; tryptain homology; v  
 C/Keywords: alternative splicing; complement classical pathway; duplication; glycoprotein  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-250/Product: complement C2b fragment #status predicted <CB2>  
 F:22-89/Domain: complement factor H repeat homology <FH1>  
 F:94-149/Domain: complement factor H repeat homology <FH2>

**Query Match**

Beat	Local Similarity	Score	IDB	Length
Match 51	Conservative	70	Mismatches	Indels Gaps

**Detailed Results:**

Cy	Db	Description
1	DLYPFLDSSGVLTTHMNIY-YFVQGLAHKFIISPOLR-NSFIYSTRTGTLMLTDERE	57
261	NLVLLDSQSSTEKDPPIPKSAELMERIFSPENVTVAILTFPASOPTIMSLERS	320
58	QIRGLELELOKLPCGDPTVMHGFEPAASEQLIYE-----NRGGYTAS----	100
321	Q-----DVTEVTLSLSDASYKDHNATGCANTVEYLIVYSMTQMOTQRDLMEISAKWEI	374
101	--VIIALTDGELEDLPFYSEBRANSHDLCAI-----VYCVGV-----KDFNE	142
375	RHTITLTIDGR--SNMGDSPKKAVTRIARELIISTEONRDVIDIALIGVKLDVDWKEINE	432
143	TOLARIADSKDHVPVNDDGFQALLGIHSIL	173
433	--LGSKDKGERHAFTLIDA-KALOQIFEHML	460

**RESULT 8**

A45974  
collagen alpha I(XIV) chain precursor, short form 2 - chicken

Nlternate names: undulin  
Cspecie: Gallus gallus (chicken)  
ClDate: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
AAccession: A45974; S30085; E22916; S17035; S20833  
RGenechr: D.R.; Foley, J.W.; Castenola, P.; Genhart, M.; Dublet, B.; Cancedda, R.; LinJ Biol. Chem. 268, 12177-12184, 1993

Title: Type XIV collagen is encoded by alternative transcripts with distinct 5' regions.

AReference number: A45974; UID:93280195; PMID:8505337

AAccession: A45974  
AStatus: Preliminary  
AMolecule type: mRNA; protein  
AResidues: 1-1747 <GER>  
ACross-references: UNIPROT:P32018  
Experimental source: embryo skin  
Note: sequence inconsistent with the nucleotide translation  
APeptide, S.B.  
submitted to the EMBL Data Library, March 1992  
AReference number: S30085  
AAccession: S30085  
AMolecule type: mRNA  
AResidues: 1472-1660 <APT>  
Cross-references: EMBL:X65122; NID:G62871; PIDN:CMA46238.1; PID:G938175  
RTrueb, J., Trueb, B.  
Eur. J. Biochem. 207, 549-557, 1992

Title: Type XIV collagen is a variant of undulin.  
Reference number: S22916; UID:9239443; PMID:1339349  
AAccession: S22916  
AStatus: preliminary  
AMolecule type: mRNA  
Residues: 286-499, 'O'; 496-834, 'A'; 836-1119, 'KU'; 1122-1402, 1409-1439 <TRU>  
R.Gordon, M.K.; Castanola, P.; Dublet, B.; Linsenmayer, T.F.; van der Rest, M.; Mayne, Eur. J. Biochem. 201, 333-338, 1991

Title: Cloning of a cDNA for a new member of the class of fibril-associated collagens  
Reference number: S17035; UID:92037585; PMID:1935930  
AAccession: S17035  
AMolecule type: mRNA  
Residues: 1472-1659 <GOR1>  
Accession: S20833

Query Match	12.2%	Score 110;	DB 2;	Length 1747;
Beet Local Similarity	26.8%;	Pred. No. 0.28;		
Matches	48;	Conservative 37;	Mismatches 74;	Indels 20; Gaps 11;
Qy	1	DLYFLIDKGSV-LHHMNEIYFYVEQL--AHKFSPO-LRMSFYSTRGTLMKLT--E	54	
Db	926	DLYFLVLDGWSLQDDNNFNKLSIFLYSTVGALDKGPQGTQVAILQFSDDPRRFFKLNAVK	985	
Qy	55	DEEOIRQGLLELOKVLPGGDTYMHGFEFASSEOIYENRQGYRTA--SVIIALTDGELHE	112	
Db	986	TKETLLEAIQOI--AVKGNTKTKGAIKHAARE-VLFGSEAGMRKGI PKVLVITDGRSOD	1042	
Qy	113	DLFFYSERANSRDL-GAIVYCVGYDPMETOLARLAD--SKOHVFPVNGFOALOGI	168	
Db	1043	DV-----NIVSRMQLDGFSPFAIGVADADYSELVNIIGSKSERVHFVVD--FDAPFTKI	1095	

```

RESULT 9
S31212
collagen alpha 1(XIV) chain precursor, short form - chicken
C/Species: Gallus gallus (chicken)
C/Date: 11-Mar-1998 #sequence_rev1501 17-Apr-1998 #text_change 15-Sep-2003
C/Accession: S31212
R/Maechli, C.; Trueb, J.; Kessler, B.; Winterhalter, K.H.; Trueb, B.
Eur. J. Biochem. 212, 483-490, 1993
A/Title: Complete primary structure of chicken collagen XIV.
A/Reference number: S31211; MUID:93185668; PMID:8444186
A/Accession: S31212
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: mRNA
A/Residues: 1-1857 <MAE>
A/Cross-references: EMBL:X70792; NID:G288874; PIDN:CA50063.1; PID:G288875
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, January 1993
C/Genetics:
A/Gene: COL14A1
C/Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F.1-28/Domain: signal sequence #status predicted <SIG>
F.29-1857/Product: collagen alpha 1(XIV) chain, short form #status predicted <MAT>
F.129-110/Domain: fibronectin type III repeat homology <FN3A>
F.156-320/Domain: von Willebrand factor type A repeat homology <VWA1>
F.352-433/Domain: fibronectin type III repeat homology <FN3B>
F.442-525/Domain: fibronectin type III repeat homology <FN3C>
F.534-614/Domain: fibronectin type III repeat homology <FN3D>
F.623-707/Domain: fibronectin type III repeat homology <FN3E>
F.741-823/Domain: F.741-823/Domain: fibronectin type III repeat homology <FN3F>
F.832-914/Domain: fibronectin type III repeat homology <FN3G>
F.922-1009/Domain: fibronectin type III repeat homology <FN3H>
F.1040-1205/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 12.2% Score 110; DB 2; Length 1857;
Best Local Similarity 26.8%; Pred. No. 0.3;
Matches 48; Conservative 37; Mismatches 74; Indels 20; Gaps 11;

1 DLYFLDKSGSV-LHHMNEYVFEQL-AHKFISPO-LRMSFVFSRGTLMKLT--E 54
|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 1042 DLYFLVDGWSGIODDNPFKIISFLYSGVGLDITGPOTGVAILQFSDDPRFPLNAK 1101

```



R.Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.  
J. Biol. Chem. 269, 20256-20262, 1994  
A>Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)  
A/Reference number: A54849; MUID:94327588; PMID:8051117  
A/Accession: A54849  
A>Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Accession: 1-2944 <CHR>  
A/Cross-references: UNIPROT:002388; GB:L02870; NID:9987124; PIDD:AAA75438.1; PID:99871220  
R.Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A>Title: Molecular cloning and characterization of type VII collagen cDNA.  
A/Reference number: P08044; MUID:922331902; PMID:1567409  
A/Accession: P08044  
A/Molecule type: mRNA  
A/Accession: 'ERR', 340-475, 'RALSTASHSTLCMRATRMHPCNCRGSHWTRAAECPCNRPASHRAAG', 524-528, 'C',  
A/Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDD:BA02853.1; PID:9453699  
A/Experimental source: keratinocyte  
A/Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
R.Partridge, M.G.; Chung, L.C.; Rymmen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Masi  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991  
A>Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A/Reference number: S16316; MUID:91334380; PMID:1871109  
A/Accession: S16316  
A/Molecule type: mRNA  
A/Accession: 'ERR', 892-892, 'E', 894-1439 <PAR>  
A/Cross-references: GB:M65156; GB:949017; NID:g180914; PIDD:AAA96439.1; PID:g180915  
A/Experimental source: keratinocyte  
R.Gammon, W.R.; Abernethy, M.L.; Pedilla, K.M.; Priyaanah, P.S.; Cook, M.E.; Wright, J.  
J. Invest. Dermatol. 99, 691-696, 1992  
A>Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protein  
A/Reference number: I56328; MUID:93107742; PMID:1465284  
A/Accession: I56328  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Accession: 'ERR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
A/Cross-references: GB:S51236; NID:g626308; PIDD:AA24637.1; PID:g626309  
R.Saltzer, U.L.; Biesen, A.Z.; Bauer, E.A.; Morris, N.P.; Granville, R.W.; Burgess, R.E.  
J. Biol. Chem. 264, 3822-3826, 1989  
A>Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
A/Reference number: A30296; MUID:89139437; PMID:2537292  
A/Accession: A30296  
A/Molecule type: protein  
A/Accession: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, 'I',  
A/Reference number: two reported peptides cannot be reliably located  
R.Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A>Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
A/Reference number: I48103; MUID:93271985; PMID:8499916  
A/Accession: I48103  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Accession: 2395-2871, 'S', 2873-2944 <RES>  
A/Cross-references: GB:L06862; NID:g388713; PIDD:AAA69196.1; PID:g388714  
R.Christiano, A.M.; Rymmen, J.; Uitto, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A>Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser substitution  
A/Reference number: A55255; MUID:94224777; PMID:8170945  
A/Accession: A55255  
A/Contents: annotation  
C/Comment: Proline and lysine at the third position of the tripeptide repeating unit (ed and subsequently O-glycosylated).  
C/Genetics:  
A/Genes: GDB:COL7A1; EBR1; EBD1; EB  
A/Cross-references: GDB:128750; OMIM:120120  
A/Map position: 3p21.3-3p21.3  
A/Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis  
A/Note: there are 118 introns  
C/Complex: type VII collagen is probably a homotrimer  
C/Description: structural component of extracellular polymer associated with anchoring  
F.11-16/Domains: coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
F.17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAR>

Query Match 11.3%; Score 102; DB 2; Length 2944;  
Best Local Similarity 26.8%; Pred. No. 2.6;  
Matches 48; Conservative 34; Mismatches 79; Indels 18; Gaps 10;

Qy 1 DLYFLIDKGSV-LHMMNEYFVEQLAHKF---ISQ-LMSFVPSRTGTT---LMLQ 52  
Db 38 DLYFLIDKSSISGRSFRERFLBGLVLPFGSAASQGVFAVQYSDDPTEFGDLAL 97

Qy 53 TEDREQIRQGLBEQKVLPCGDDTYMHGFERASQIYYENRQGYRTAIVITALTQGLHR 112  
Db 98 GGGGVDIR-AIRELS--YKGNTRTGAIILHVDHVLPOLARQGVPCVILITDCK-SQ 153

Qy 113 DLFYSERANRSDICAIYVCVGDQFNQTOLARIAD--SKDHVPVNDGFOALQGI 169  
Db 154 DLV---DTAQRKGGGVKLFVAGIKVADPELKRVASQPSDPFFVND-FSLIRTL 208

RESULT 14

BBHU  
Complement factor B precursor [validated] - human  
N:Alternate names: C3 convertase; C3 proactivator; glycine-rich beta-glycoprotein; heat-  
N:Contains: alternative-complement-pathway C3/C5 convertase (EC 3.4.21.47) Bb fragment  
C:Species: Homo sapiens (man)  
C:Date: 19-Feb-1984 #sequence, revision 05-Aug-1994 #text change 09-Jul-2004  
C:Accession: S34075; A44622; A00934; A19188; A19947; B19947; S14339; A44628; I544  
R:Mejia, J.R.; Jahn, I.; de la Salle, H.; Hauptmann, G.  
submitted to the EMBL Data Library, March 1993

A:Reference number: S34075  
A:Accession: S34075  
A:Molecule type: mRNA  
A:Residues: 1-764 <MEJ>  
A:Cross-references: UNIPROT:P00751; EMBL:X72875; NID:g297568; PIDN:CA51389.1; PID:g29755  
R:Moore, D.R.; Marham, A.F.; Richer, A.T.; Goldeberger, G.; Colten, H.R.  
Proc. Natl. Acad. Sci. U.S.A. 79, 5661-5665, 1982  
A:Title: Isolation of cDNA clones for the human complement protein factor B, a class III  
A:Reference number: A44622; MOID:83039428; PMID:6957884  
A:Accession: A44622  
A:Molecule type: mRNA  
A:Residues: 467-546;550-595;752-764 <MOO>  
A:Cross-references: GB:J00185; GB:J00186  
R:Moore, J.E.; Anderson, J.K.; Davison, B.A.; Woods, D.E.  
J. Biol. Chem. 259, 3407-3412, 1984  
A:Title: Complete primary structure for the zymogen of human complement factor B.  
A:Reference number: A20751; MOID:84161997; PMID:6546754  
A:Accession: A00934  
A:Molecule type: protein; mRNA

A/Residues: 26-764 <MOL>  
A/Cross-references: GB:K01566  
A/Note: nucleic acid translation differs from the sequence shown in having 300-Leu, 328-  
A/Note: 736-Ser was also found  
A/Note: glycosylation sites were determined  
R/Character: D.L.; Gagnon, J  
Biochem. J. 209, 61-70, 1983  
A/Title: Amino acid sequence of the Bb fragment from complement factor B. Sequence of th  
A/Reference number: A19188; MUID:83204002; PMID:6342610  
A/Content: the final paper in a series documenting the sequence, glycosylation site, an  
A/Accession: A19188  
A/Molecule type: protein  
A/Residues: 260-236, 'T', 298-764 <CHR>  
R/Campbell, R.D.; Porter, R.R.  
Proc. Natl. Acad. Sci. U.S.A. 80, 4464-4468, 1983  
A/Title: Molecular cloning and characterization of the gene coding for human complement  
A/Reference number: A19947; MUID:83272641; PMID:6308626  
A/Accession: A19947  
A/Molecule type: DNA  
A/Residues: 346-764 <CAM>  
A/Cross-references: GB:U00125  
A/Accession: B19947  
A/Molecule type: mRNA  
A/Residues: 339-509 <CA1>  
A/Cross-references: GB:U00126; NID:g187723; PIDN:AAA36226.1; PID:g553536  
R/Wu, L.; Morley, B.J.; Campbell, R.D.  
Cell 48, 331-342, 1987  
A/Title: Cell-specific expression of the human complement protein factor B gene: evidenc  
A/Reference number: A25971; MUID:87102880; PMID:3643061  
A/Accession: B25971  
A/Molecule type: DNA  
A/Residues: 1-99 <MOL>  
A/Cross-references: GB:M15082; NID:g187699; PIDN:AAA59625.1; PID:g553534  
R/Niemann, M.A.; Bhown, A.S.; Miller, E.J.  
Biochem. J. 274, 473-480, 1991  
A/Title: The principal site of glycation of human complement Factor B.  
A/Reference number: S14339; MUID:91174758; PMID:2006911  
A/Accession: S14339  
A/Molecule type: protein  
A/Residues: 270-329 <NIE>  
A/Note: binding site for carbohydrate to lysine under artificial conditions  
R/Morley, B.J.; Campbell, R.D.  
EMBO J. 3, 153-157, 1984  
A/Title: Internal homologues of the Ba fragment from human complement component factor B  
A/Reference number: A44628; MUID:84158524; PMID:6323161  
A/Accession: A44628  
A/Molecule type: preliminary  
A/Status: preliminary  
A/Residues: 16-225, 'F', 227-259 <MOR>  
R/Schwedde, W.; Lutrig, B.; Sokolowski, T.; Eschaller, C.; Weiss, E.H.; Meyer zum Busche  
Immunobiology 188, 221-232, 1993  
A/Title: Human complement factor B: functional properties of a recombinant zymogen of th  
A/Reference number: I54409; MUID:94041399; PMID:8225386  
A/Accession: I54409  
A/Status: translated from GB/EMBL/DBDJB  
A/Molecule type: mRNA  
A/Residues: 1-764 <RES>  
A/Cross-references: GB:S67310; NID:g452937; PIDN:AA13989.1; PID:g4261689  
R/Hortlich, T.; Klm, S.; Matsumoto, M.; Watanabe, I.; Fujita, S.; Volanakis, J.E.  
Mol. Immunol. 30, 1587-1592, 1993  
A/Title: Human complement factor B: cDNA cloning, nucleotide sequencing, phenotypic cony  
A/Reference number: I57824; MUID:94061177; PMID:8247029  
A/Accession: I57824  
A/Status: translated from GB/EMBL/DBDJB  
A/Molecule type: mRNA  
A/Residues: 1-31, 'O', 33-764 <RE2>  
A/Cross-references: GB:L15702; NID:g291921; PIDN:AAA16820.1; PID:g291922  
C/Comment: 292-Cys has a free sulphydryl.  
C/Genetics:  
A/Gene: GDB:BF  
A/Cross-references: GDB:119726; OMIM:138470  
A/Map position: 6p21.3-6p21.3  
A/Intons: 21/3; 99/3; 346/1; 390/1; 424/1; 470/1; 502/3; 542/1; 593/2; 619/1; 652/3; 69

A/Note: the list of introns may be incomplete  
A/Note: gene is located in the major histocompatibility complex, class III region  
A/Complex: complement factor B initially forms an inactive complex with complement factor  
ment factor C3b forming active C3/C5 convertase; Ba is released  
C/Function:  
A/Description: Bb is a serine protease; C3/C5 convertase cleaves complement C3 alpha ct  
a1  
A/Pathway: complement alternate pathway  
C/Superfamily: complement B/C2; complement factor H repeat homology; trypsin homology; v  
C/Keywords: acute phase; complement alternate pathway; duplication; glycoprotein; hydroly  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-764/Product: complement factor B #status experimental <MAT>  
F:26-259/Product: complement factor Ba fragment #status experimental <BAF>  
F:37-98/Domain: complement factor H repeat homology <PH1>  
F:103-158/Domain: complement factor H repeat homology <PH2>  
F:165-218/Domain: complement factor H repeat homology <PH3>  
F:260-764/Product: C3/C5 convertase Bb fragment #status experimental <BBF>  
F:268-458/Domain: von Willebrand factor type A repeat homology <VFA>  
F:482-752/Domain: trypsin homology #status atypical <TRY>  
F:37-76, 62-98, 103-145, 131-158, 165-205, 191-218, 478-596, 511-527, 599-615, 656-682, 695-725/Dic  
F:122, 142, 285, 378/Binding site: carbohydrate (asn) (covalent) #status experimental  
F:259-260/Cleavage site: Arg-Lys (complement factor D) #status experimental  
F:526, 576, 699/Active site: His, Asp, Ser #status experimental

Query Match 11.3%; Score 101.5; DB 1; Length 764;  
Best Local Similarity 22.5%; Pred. No. 0.59;  
Matches 47; Conservative 43; Mismatches 74; Indels 45; Gaps 10;

QY 1 DLYFIIDKSGSV-----LHHMNEIYFVBOIAKFIISPOLRMSFIYSTRGTTLMKTE- 54  
DB 270 NITVLDDGSDISANSFTGAKCLVNIIEKVASGYKVP--RYGLVYATIPKVIWAKVSEA 327  
QY 55 ---DREQIRQGLBEL-----OKVLPGDPTVYHGESEFASQIYENR-----QGY-RTA 99  
DB 328 DSNNAWVTFKQLMEIYVEDHKLSGNT-----KALQAVYSWMSWPDVPEGNVTR 381  
QY 100 SYVIALTDG-----ELHEDLPFYSREANRSDLGAIYCVG--VKDFNETQ 144  
DB 382 HVIILMTDGLHNMGDPITVIDEIRDLVIGKGRKPRREDYLDVYFVGVLVQVINA 441  
QY 145 LARIADSKDHPVFNDFQALGIIHSIL 173  
DB 442 LASKDNEQHFVKVD-MENTLEDVYQWI 469

RESULT 15  
T28797  
hypothetical protein C16B9.1 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T28797  
R/Geisels, C.  
submitted to the EMBL Data Library, November 1995  
A/Description: The sequence of C. elegans cosmid C16B9.  
A/Reference number: Z20525  
A/Accession: T28797  
A/Status: preliminary; translated from GB/EMBL/DBDJB  
A/Molecule type: DNA  
A/Residues: 1-567 <GB1>  
A/Cross-references: UNIPROT:Q18048; EMBL:U09677; PIDN:AA47957.1; GSPDB:GN00028; CESP:CI  
A/Experimental source: strain Bristol N2; clone C16B9  
C/Genetics:  
A/Gene: CESP:C16B9.1  
A/Map position: X  
A/Intons: 15/3; 77/3; 132/1; 185/3; 231/1; 271/2; 301/2; 430/3; 509/1

Query Match 11.2%; Score 100.5; DB 2; Length 567;  
Best Local Similarity 24.1%; Pred. No. 0.5;  
Matches 41; Conservative 38; Mismatches 72; Indels 19; Gaps 9;

QY 1 DLYFIIDKSGSVLHHMNEIYFVBOIAKFIISPOLRMSFIYSTRGTTLMKTE 56  
DB 393 DIGIIFDSSGSLKKNFQKQALVBOQPISDNATRGIVGFAGKTKRVLANFSQNK 452



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 13, 2005, 19:37:47 ; Search time 58.0753 Seconds  
(without alignments)  
1525.429 Million cell updates/sec

Title: US-09-970-076-2\_COPY\_44\_216  
Perfect score: 899  
Sequence: 1 DLYFILDKSGSVLHMNEIY.....HVFVNDGFGALQGIHSIL 173

Scoring table: BLASTUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seegs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03.\*  
1: uniProt\_sprot.\*  
2: uniProt\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	899	100.0	564	1 ATRI_HUMAN	Q96X22 mus sapien
2	890	99.0	562	1 ATRI_MOUSE	Q96X22 mus musculu
3	519	57.7	489	1 ATRI_HUMAN	P58335 mus sapien
4	517	57.5	487	2 O6DFX2	O6DFX2 mus musculu
5	348	38.7	641	2 Q8BVW2	Q8BVW2 mus musculu
6	141	15.7	1332	2 Q9BPQ8	Q9BPQ8 halocynthia
7	138.5	15.4	1161	1 ITAD_RAT	Q96Y67 rattus norv
8	129.5	14.4	724	2 O04588	O04588 elmeria max
9	128.5	14.3	1152	1 ITAM_HUMAN	P11215 mus sapien
10	128.5	14.3	3119	1 CAIC_MOUSE	O60847 mus musculu
11	125	13.9	441	2 O8T6T5	O8T6T5 mytilus edu
12	125	13.9	444	2 O8T6T5	O8T6T5 mytilus edu
13	124	13.8	3124	1 CAIC_CHICK	P13944 gallus galli
14	124	13.8	765	2 Q9U8J9	Q9U8J9 neospora ca
15	122.5	13.6	1162	1 ITAD_HUMAN	O13349 mus sapien
16	122	13.6	453	2 O8T5C2	O8T5C2 mytilus gal
17	121.5	13.5	712	2 O43981	O43981 elmeria ten
18	121.5	13.5	1163	1 ITAM_HUMAN	P20702 mus sapien
19	120.5	13.4	1153	1 ITAM_MOUSE	P05555 mus musculu
20	118.5	13.2	3063	1 CAIC_HUMAN	Q99715 mus sapien
21	117	13.0	1182	2 O8C6K9	O8C6K9 mus musculu
22	117	13.0	1823	2 O7PRP5	O7PRP5 anopheles g
23	116.5	13.0	760	2 O70350	O70350 mus musculu
24	116.5	13.0	920	2 Q28984	Q28984 sus scrofa
25	116	12.9	637	2 O81VX1	O81VX1 mus sapien
26	116	12.9	1284	2 O6P159	O6P159 mus sapien
27	116	12.9	1329	1 KPI10_HUMAN	Q99218 mus sapien
28	113.5	12.6	517	2 O43853	O43853 mus sapien
29	113.5	12.6	760	1 CO2_MOUSE	P21180 mus musculu
30	112	12.5	588	2 O6ZED8	O6ZED8 synecocyst
31	111.5	12.4	2104	2 Q21281	Q21281 caenorhabdi

32	111.5	12.4	2104	2 Q964N4	Q964N4 caenorhabdi
33	111	12.3	440	2 Q8CBT2	Q8CBT2 mus musculu
34	111	12.3	848	2 Q8C720	Q8C720 mus musculu
35	111	12.3	3567	2 Q9ES77	Q9ES77 mus sapien
36	110	12.2	1626	2 O8NFW1	O8NFW1 mus sapien
37	110	12.2	1888	1 CAIC_CHICK	P12018 gallus galli
38	109	12.1	3183	2 O65ZC2	O65ZC2 caenorhabdi
39	109	12.1	3767	1 M0A3_CAEEL	P34576 caenorhabdi
40	108	12.0	1188	1 ITAM_MOUSE	P61622 mus musculu
41	108	12.0	1188	2 Q7TQC3	Q7TQC3 mus musculu
42	107.5	12.0	599	2 O8MVQ1	O8MVQ1 bolitena vi
43	107	11.9	919	2 O7SR52	O7SR52 lymnaea ata
44	107	11.9	2944	2 O63870	O63870 mus musculu
45	106.5	11.8	415	2 Q8DGY6	Q8DGY6 synecococc

## ALIGNMENTS

RESULT 1  
ID ATRI\_HUMAN STANDARD; PRT; 564 AA.  
AC Q96X22; Q9NVP3;  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
GN Name=ANTXR1; Synonyms=ATR, TEM8;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCB1\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20407466; PubMed=10947988; DOI=10.1126/science.289.5482.1197;  
RA St Croix B., Rago C., Velculescu V., Traverso G., Romans K.B.,  
RA Montgomery E., Lal A., Riggs G.J., Lengauer G., Vogelstein B.,  
RA Kinzler K.W.;  
RT "Genes expressed in human tumor endothelium";  
RL Science 289:1197-1202(2000).  
[2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH ANTHRAX TOXIN.  
RX MEDLINE=21557240; PubMed=11700562; DOI=10.1038/n510198;  
RA Bradley K.A., Mogridge J., Mourez M., Collier R.J., Young J.A.T.;  
RL "Identification of the cellular receptor for anthrax toxin.";  
RL Nature 414:225-229(2001).  
[3]  
RP SEQUENCE OF 184-564 FROM N.A. (ISOFORM 1), AND SEQUENCE FROM N.A.  
(ISOFORM 3).  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakematsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohtsushi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,  
RA Nagahashi K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,  
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Komoto H.,  
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,  
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,  
RA Yamazaki M., Nishimura K., Iehibashi T., Yamashita H., Murakawa K.,  
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y.,  
RA Iehida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotsu T.,  
RA Kuwano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,  
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,  
RA Imose N., Mushino K., Yuki H., Ohshima A., Sasaki N., Aotaka S.,  
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Nomiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Hishigaki S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamaguchi M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
RA Kawabata A., Hikiyama T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,

RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togaishi T., Oyama M., Hata H., Watanabe Y., Komatsu T.,  
 RA Mitushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuo Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.,  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs";  
 RL Nat. Genet. 36:40-45(2004).  
 RN [4]  
 RN SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strauberg R.L., Fejngold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Shermen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stempleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinci P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RN INTERACTION WITH ANTHRAX TOXIN.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;  
 RA Scobie H.M., Ratney G.J.A., Bradley K.A., Young J.A.T.,  
 RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin  
 RT receptor";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).  
 RN [6]  
 RN SPICE ISOFORM(S) THAT ARE POTENTIAL NMD TARGET(S).  
 RX PubMed=14759258; DOI=10.1186/gb-2004-5-2-r8;  
 RA Hallman R.T., Green R.E., Brenner S.E.,  
 RT "An unappreciated role for RNA surveillance";  
 RL Genome Biol. 5:RESEARCH008.1-RESEARCH008.16(2004).  
 CC -1- FUNCTION: Cellular role is not yet known.  
 CC -1- SUBUNIT: Binds to the protective antigen (PA) of *Bacillus*  
 CC anthracis. Binding does not occur in the presence of calcium.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Comment=Experimental confirmation may be lacking for some  
 CC isoforms;  
 CC Name=1;  
 CC IsoId=Q9H6X2-1; Sequence=Displayed;  
 CC Note=May be produced at very low levels due to a premature stop  
 CC codon in the mRNA, leading to nonsense-mediated mRNA decay;  
 CC Name=2;  
 CC IsoId=Q9H6X2-2; Sequence=VSP\_000444; VSP\_000445;  
 CC Name=3;  
 CC IsoId=Q9H6X2-3; Sequence=VSP\_000446; VSP\_000447;  
 CC Name=4;  
 CC IsoId=Q9H6X2-4; Sequence=VSP\_000448; VSP\_000449;  
 CC -1- TISSUE SPECIFICITY: Highly expressed in tumor endothelial cells  
 CC but not in normal endothelial cells.  
 CC -1- DOMAIN: Binding to PA seems to be effected through the VMA domain.  
 CC -1- SIMILARITY: Belongs to the ATR family.  
 CC -1- SIMILARITY: Contains 1 VMA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: AF279145; AAK52094.1; -  
 DR EMBL: AF421380; BAI26496.1; -  
 DR EMBL: AK025429; BAB15128.1; ALT\_INIT.  
 DR EMBL: AK001463; BAA91707.1; ALT\_FRAME.  
 DR EMBL: BC012074; AAA12074.1; -  
 DR Genew: HGNC:21014; ANTXR1.  
 DR H-InvDB: HIX0002125; -  
 DR MIM: 606410; -  
 DR InterPro: IPR008400; Anth\_Ig.  
 DR InterPro: IPR008399; Ant\_C.  
 DR InterPro: IPR02035; VWF\_A.  
 DR Pfam: PF05587; Anth\_Ig\_1.  
 DR Pfam: PF05566; Ant\_C\_1.  
 DR Pfam: PF00092; VMA\_1.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PS50234; VWF\_A; 1.  
 DR Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.  
 FT SIGNAL 1 32  
 FT CHAIN 1 32  
 FT DOMAIN 33 321  
 FT TRANSMEM 322 342  
 FT DOMAIN 343 564  
 FT DOMAIN 44 215  
 FT DOMAIN 360 368  
 FT DOMAIN 506 564  
 FT CARBOHYD 166 166  
 FT CARBOHYD 184 184  
 FT CARBOHYD 262 262  
 FT VARSPPLIC 365 368  
 FT VARSPPLIC 369 564  
 FT VARSPPLIC 268 297  
 FT VARSPPLIC 298 564  
 FT VARSPPLIC 319 333  
 FT VARSPPLIC 334 564  
 FT SEQUENCE 564 AA; 62789 MW; B118A00D5DF2233 CRC64;  
 Query Match 100.0%; Score 899; DB 1; Length 564;  
 Best Local Similarity 100.0%; Pred. No. 1e-69;  
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DLVFIIDKSGSVLHHNNEYFYFEQLAHKFIISQLMSFVSTRTGTLTKLLEDPKQIR 60  
 DB 44 DLVFIIDKSGSVLHHNNEYFYFEQLAHKFIISQLMSFVSTRTGTLTKLLEDPKQIR 103  
 QY 61 QGLEBEIQKLVPGSDTYMHSEGFERASQIYYENQGYRTASVIALTDGELHEDLPFYSER 120  
 DB 104 QGLEBEIQKLVPGSDTYMHSEGFERASQIYYENQGYRTASVIALTDGELHEDLPFYSER 163  
 QY 121 EANRSRDLAGIYVCVGVKDFNETQIARIADSKDHVPVNDGFOALQGIHSIL 173  
 DB 164 EANRSRDLAGIYVCVGVKDFNETQIARIADSKDHVPVNDGFOALQGIHSIL 216  
 RESULT 2  
 ATRI\_MOUSE  
 ID ATRI\_MOUSE STANDARD; PRT; 562 AA.  
 AC Q9C252;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Anthrax toxin receptor 1 precursor (Tumor endothelial marker 8).  
 CN Name=Anthrax1, Synonyms=Atcr, Tem8;  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 [1]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP MEDLINE=21443268; PubMed=11559528;  
 RX Carson-Walter E.B., Watkins D.N., Nanda A., Vogelstein B.,  
 RA Kinzler K.W., St Croix B.;  
 RT "Cell surface tumor endothelial markers are conserved in mice and  
 humane".  
 RL Cancer Res. 61:6649-6655(2001).  
 [2]  
 RN SEQUENCE OF 88-562 FROM N.A. (ISOFORM 2).  
 RP STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=22354683; PubMed=12466651; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Mikaido I., Oeato N., Saito R., Suzuki H., Yamana H., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Balderis R.L., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schmitt L.M., Knapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Brad D., Brusic V., Chochia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dregani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Gilmord S., Guelinowski S., Hirokawa N., Jackson I.J., Jarvis B.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Natta K., Okido T., Pavan W.J., Petrea G., Pebole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravaei T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Vervard R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wyshew-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuen Z., Zavalan M., Zhu Y., Zimmer A., Carlini P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arikawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imocani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs".  
 RL Nature 420:563-573(2002).  
 CC -1- FUNCTION: Cellular role is not yet known.  
 CC -1- SUBUNIT: Binds to the protective antigen (PA) of Bacillus  
 CC anthracis (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O9C252-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O9C252-2; Sequence=VSP\_000450;  
 CC Note=No experimental confirmation available;  
 CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain  
 CC (By similarity).  
 CC -1- SIMILARITY: Belongs to the ATR family.  
 CC -1- SIMILARITY: Contains 1 VWA domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on free  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL, AF378762; AAL1199.1; -;  
 DR EMBL, AK013005; BAB2859.1; -; ALT INIT.

DR	MGI:1916788; Antxrl.
DR	InterPro; IPR008460; Antch_Ig.
DR	InterPro; IPR008399; Ant_C.
DR	InterPro; IPR002035; VWF_A.
DR	Pfam; PF05587; Antch_Ig; 1.
DR	Pfam; PF05586; Ant_C; 1.
DR	Pfam; PF00092; VWF_1.
DR	PROSITE; PS50234; VWF_A; 1.
KW	Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.
FT	SIGNAL
FT	CHAIN
FT	DOMAIN
FT	TRANSMEM
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	CARBOHYD
FT	CARBOHYD
FT	CARBOHYD
FT	VARSPLIC
FT	SEQUENCE
FT	Best Local Similarity
FT	Matches
OY	1 DLYPILDKSGSVLHNMNIYFVEQLAHKFIISPOLMSFIYSRGTTLMLKTEDREOR 60
DB	42 DLYPILDKSGSVLHNMNIYFVEQLAHKFIISPOLMSFIYSRGTTLMLKTEDREOR 101
OY	61 QGSELEQVLPAGCTVMHEGEFRASEQIYYNRCGYTASTITLTGSELHEDIFFYSER 120
DB	102 QGSELEQVLPAGSDTYMHEGEFRASEQIYYNSQGRASYITALTGELHEDIFFYSER 161
OY	121 EANSRPDLGAIVYCVGVDPFNETOLARIADSKDHVPVNDGFQALQGIIHSIL 173
DB	162 EANSRPDLGAIVYCVGVDPFNETOLARIADSKDHVPVNDGFQALQGIIHSIL 214
RESULT 3	
ATRX2_HUMAN STANDARD; PRT; 489 AA.	
ID	_ATRX2_HUMAN
AC	P58335; Q8BUI1; Q8NB13; Q96NC7;
DT	28-FEB-2003 (Rel. 41, Created)
DT	10-OCT-2003 (Rel. 42, Last sequence update)
DT	05-JUL-2004 (Rel. 44, Last annotation update)
DE	Anthrax toxin receptor 2 precursor (Capillary morphogenesis protein-2) (CMC-2).
GN	Name=ANTXR2; Synonym=CDC2;
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RX	SEQUENCE FROM N.A. (ISOFORM 2).
RX	MEDIUN=21539596; PubMed=11683410;
RA	Bell S.E., Mavila A., Salazar R., Bayless K.J., Kanagala S.,
RA	Maxwell S.A., Davis G.E.;
RT	"differential gene expression during capillary morphogenesis in 3D
RT	collagen matrices: regulated expression of genes involved in basement
RT	membrane matrix assembly, cell cycle progression, cellular
RT	differentiation and G-protein signaling.";
RL	J. Cell Sci. 114:2755-2773(2001).
RP	[2]
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH ANTHRAX TOXIN.
RC	TISSUE=Placenta;
RX	MEDIUN=22608610; PubMed=12700348; DOI=10.1073/pnas.0431098100;
RA	Scobie H.M., Rainey G.J.A., Bradley K.A., Young J.A.T.;

RT "Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor." Proc. Natl. Acad. Sci. U.S.A. 100:5170-5174(2003).

RN [3]

RP SEQUENCE OF 78-489 FROM N.A. (ISOFORM 3), AND SEQUENCE FROM N.A. (ISOFORM 4).

RC TISSUE=Synovial cell;

RX PubMed14702039; DOI=10.1038/ng1285;

RA Ota T., Suzuki Y., Nishikawa T., Ohtsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Nimomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori Y., Tanai H., Kimata M., Watanabe S., Yoshida M., Horita T., Iuhida S., Ono Y., Takaguchi S., Watanabe S., Yoshida M., Horita T., Kusanagi J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Ooshima A., Sasaki N., Aotsuka S., Yoshihara Y., Matsunawa H., Ichihara T., Shiohara N., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Moriya S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiyama T., Oho T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T., Mizushima-Shigino K., Sato T., Shirai Y., Takahashi Y., Nakagawa K., Mizushima K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isozaki T., Sugano S.; "Complete sequencing and characterization of 21,243 full-length human cDNAs." Nat. Genet. 36:40-45(2004).

RL Nat. Genet. 36:40-45(2004).

RT "FUNCTION: Cellular role is not yet known.

CC -1- SUBUNIT: Binds to the protective antigen (PA) of *Bacillus anthracis* in a divalent cation-dependent manner, with the following preference: calcium > manganese > magnesium > zinc.

CC Seems to bind to collagen type IV and laminin.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1 and 2). Secreted (isoform 3). Isoform 1 is expressed at the cell surface while isoform 2 is predominantly expressed within the endoplasmic reticulum and not at the plasma membrane.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=4;

CC Name=1;

CC IsoId=P58335-1; Sequence=Displayed;

CC Name=2;

CC IsoId=P58335-2; Sequence=VSP\_008343;

CC Name=3;

CC IsoId=P58335-3; Sequence=VSP\_008344, VSP\_008345;

CC Name=4;

CC Note=No experimental confirmation available;

CC Name=4;

CC IsoId=P58335-4; Sequence=VSP\_008346;

CC Note=No experimental confirmation available;

CC -1- TISSUE SPECIFICITY: Expressed in colon, heart, kidney, lung, liver, peripheral blood leukocytes, placenta, skeletal muscle, small intestine and spleen.

CC -1- DOMAIN: Binding to PA seems to be effected through the VWA domain.

CC -1- SIMILARITY: Belongs to the ATR family.

CC -1- SIMILARITY: Contains 1 VWA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL: AY040326; AAK7722.1; -.

DR EMBL: AY233452; AAP04016.1; -.

DR EMBL: AK055636; BAB70976.1; ALT\_INT.

DR EMBL: AK091721; BAC03731.1; -.

DR GenBank: HGNC:21732; ANTXR2.

DR MIM: 608041; -.

DR InterPro: IPR008399; Ant\_C.

DR InterPro: IPR02035; VWF\_A.

DR Pfam: PF05586; Ant\_C; 1.

DR Pfam: PF00092; VWA\_1.

DR PROSITE: PSS0234; VWF, 1.

KM Alternative splicing; Glycoprotein; Receptor; Signal; Transmembrane.

FT SIGNAL 1 33

FT CHAIN 34 489

FT DOMAIN 34 318

FT TRANSMEM 319 341

FT DOMAIN 342 489

FT DOMAIN 44 213

FT CARBOHYD 250 250

FT CARBOHYD 260 260

FT VARSPPLIC 213 315

FT VARSPPLIC 290 322

FT VARSPPLIC 323 489

FT VARSPPLIC 477 489

FT CONFLICT 357 357

FT SEQUENCE 489 AA; 53692 MW; B9F679DB75B6E2B7 CRC64;

Query Match 57.7%; Score 519; DB 1; Length 489;

Best Local Similarity 59.5%; Pred. No. 9.7e-37;

Matches 103; Conservativity 34; Mismatches 34; Indels 2; Gaps 1;

QY 1 DLVFIIDKSGSVVHHNNIEYFVEQLAHKRPISQOLNMFVFSRGTGLMKLTFRDQIR 60

DB 44 DLFIIVDKSGSVANNMIEIYFVQLAERFVSQEMKLSFVSQATITIIPLTGDRKIS 103

QY 61 QGLEELQKLVPGGDYTHHEGFEASQIYYENQGYRTASVIAITLDEGLHEDLFFYSER 120

DB 104 KGLEELKRVSPVGETYTHHEGLKLANQI--QKAGGLKTSIIITLDTGKLDGLVPSYAEK 161

QY 121 EAKRSRDLGAIYVCVGVKDPNEFQRLARIADSKRHVPVNDGFQALGIIISIL 173

DB 162 EAKISRSLSGASVYCVGLDFEQALERIADSKQVFPVKGQFQALGIIISIL 214

RESULT 4

Q6DFX2

ID Q6DFX2 PRELIMINARY; PRT; 487 AA.

AC Q6DFX2;

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Anthrax toxin receptor 2.

GN Name=Antxr2;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Eye; DOI=10.1073/pnas.242603899;

RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold B.A., Grouse L.H., Dege J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hatleh P.,  
 RA Dlatcenco L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshitsuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Boesak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Holyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fehey J., Helton E., Kottman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzyzanski M.I., Skelton U., Smalhus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Eye;  
 RA Strauberg R.;  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC076595; AAH76595.1; -  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0004872; F:receptor activity; IEA.  
 DR Interpro: IPR008400; AnCh\_IG.  
 DR Interpro: IPR008399; AnCh\_IG.  
 DR Interpro: IPR02035; VWF\_A.  
 DR Pfam: PF05587; AnCh\_IG; I.  
 DR Pfam: PF05586; AnCh\_IG; I.  
 DR Pfam: PF00092; VMA; 1.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PS50234; VWF\_A; 1.  
 KM Receptor.  
 SQ SEQUENCE 487 AA; 53184 MW; 61A400D60BCD5E9 CRC64;  
 Query Match 57.5%; Score 517; DB 2; Length 487;  
 Best Local Similarity 59.5%; Pred. No. 1.4e-36;  
 Matches 103; Conservative 31; Mismatches 37; Indels 2; Gaps 1;  
 QY 1 DLYPIIDKSGSVLHNNIEIYFVEQLAHKFIISPOLKMSFIVFSTRTGTLTKLTEDRQIR 60  
 DB 44 DLYFVLDKSGSVANNWIEIYNFVHQLTERFVSPSPMRISFIVSSQATIIIPLTGDRYKIG 103  
 QY 61 OGLEELQKLVPGCGDTYHGEFPERASBOIYYENRQGYRTASYIILTDGELHEDLPFYSER 120  
 DB 104 KGEEDLKAVFVGETIYHGEKLANEQI--QNGAGLKASIIITLTGKLDGLVPSYAE 161  
 QY 121 EANRSRDGAIYVCVGVKDFNETQLARIADSKCHVFPVNDGFPALQGIHSITL 173  
 DB 162 EAKKSRELASVYCVGVLDPEQQLERLADSKDQVFPVKGQFQLKGIINSITL 214  
 RESULT 5  
 ID Q8BVM2 PRELIMINARY; PRT; 641 AA.  
 AC Q8BVM2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Mus musculus adult male testis cDNA, RIKEN full-length enriched  
 DE library, clone:933430J1 product:hypothetical Prolin-rich region/von  
 DE Willebrand factor type A domain containing protein, full insert  
 DE sequence.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;

RT "High-efficiency full-length cDNA cloning.";  
 RL Mech. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subcloning of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=20530913; PubMed=11078861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitanaka T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaibinagi K.,  
 RA Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsubara S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasuawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akaiura S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AK077206; BAC36683.1; -  
 DR GO: GO:0016021; C:integral to membrane; IEA.  
 DR GO: GO:0004872; F:receptor activity; IEA.  
 DR Interpro: IPR008400; AnCh\_IG.  
 DR Interpro: IPR002035; VWF\_A.  
 DR Pfam: PF05587; AnCh\_IG; I.  
 DR Pfam: PF00092; VMA; 1.  
 DR PRINTS: PR00453; VWFADOMAIN.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PS50234; VWF\_A; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 641 AA; 70415 MW; 199B300730B85B3 CRC64;  
 Query Match 38.7%; Score 348; DB 2; Length 641;  
 Best Local Similarity 43.8%; Pred. No. 9.8e-22;  
 Matches 74; Conservative 30; Mismatches 65; Indels 0; Gaps 0;  
 QY 1 DLYPIIDKSGSVLHNNIEIYFVEQLAHKFIISPOLKMSFIVFSTRTGTLTKLTEDRQIR 60

```

Db      76 DLYVLDKSSGVANWNIHITSPAGLVKFTNPNLRISITITSEAVIITLSDSKEN 135
Qy      61 QGLEELKVLPGDPTWMEGPERASEQIYENROGYRTASVIALTDGELHEDLPFYSER 120
Db      136 KSLIVLSIVPGGLTHMQKLRKANEGIRKSTLGGRIANSVIALTDGLLLKRYLPTME 195
Qy      121 EANSRDLGAIYCVGVKDFMETOLARADSKHVPFNPNGFOLQGIT 169
Db      196 EAKKARBMGAIVYGVFMYSKQQLVNIAGDPDRCFVDEGFSALGVV 244

RESULT 6
Q9BP08 ID Q9BP08 PRELIMINARY; PRT; 1332 AA.
AC Q9BP08 01-JUN-2001 (TREMBLERL 17, Created)
DT 01-JUN-2001 (TREMBLERL 17, Last sequence update)
DT 01-JUN-2003 (TREMBLERL 24, Last annotation update)
DE Integrin alpha Hrl precursor.
GN Name=HrTGA1;
OS Halocynthia roretzi (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Scolidobranchia; Pyuridae; Halocynthia.
OX NCBI_TaxID=7729;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hemocyte;
RC MEDLINE=21103187; PubMed=11160215;
RA Miyazawa S., Azumi K., Nonaka M.;
RT "Cloning and characterization of integrin alpha subunits from the
RT solitary ascidian, Halocynthia roretzi."
RL J. Immunol. 166:1710-1715(2001).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
DR EMBL; AB048261; BAB21479.1; -.
DR HSSP; P11215; IBHO.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0005515; P:protein binding; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR GO; GO:0007229; P:integrin-mediated signaling pathway; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VWA; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR SMART; SM00453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
DR Cell adhesion; Integrin; Signal; Transmembrane.
FT SIGNAL 31 1332 potential.
FT CHAIN 1 1332 integrin alpha Hrl.
SQ SEQUENCE 1332 AA; 145851 MW; 0D9108D2B05CFEAB CRC64;

Query Match 15.7%; Score 141; DB 2; Length 1332;
Best Local Similarity 24.3%; Pred. No. 0.0023;
Matches 51; Conservative 34; Mismatches 59; Indels 66; Gaps 7;

Qy      1 DLYFILDKSSGVHNMNIYFVQLAKFISPLRMSFYFSRGTTLTKMLTDRDQIR 60
Db      206 DVLFLVLDKSSGVGNPKFVKDWNKNT-----AKLDIGKEIVR 243
Qy      61 QGLEELKVLPGDPTWMEGPERASEQIYENROGYRT----- 98
Db      244 VGVVQVSHYVEGKINKOKYITTEISIGFGLDLPFNADVRI---QLQGYTYTGRALQ 300
Qy      99 -----ASVIALTDGELHEDLPFYSERANRSRDGAIYCVGVKDFMETOL 145
Db      301 KVIRDFPDAYIGNKQVALLLTLDGAKDNKILP--NANRLNKGIATFAVGVEYDISEL 358
Qy      146 ARIA---DSKDHPFVNDGFOLQGITHSI 172

```

```

Db      359 KLIASGDSIDRYFTYTD--FGELDSIVKSL 387

RESULT 7
ITAD RAT ID ITAD RAT STANDARD; PRT; 1161 AA.
AC Q9OYE7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Integrin alpha-D precursor.
GN Name=Itgad;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RC O'Brien M.M., VanderVlieten M., Kilgannon P.D., Dietrich G.,
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin."
RL Submitted (JUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and
CC VCAM1. May play a role in the atherosclerotic process such as
CC clearing lipoproteins from plaques and in phagocytosis of blood-
CC borne pathogens, particulate matter, and senescent erythrocytes
CC from the blood (By similarity).
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D
CC associates with beta-2 (By similarity).
CC -1- DOMAIN: The integrin I-domain (Inert) is a VWF_A domain. Integrins
CC with I-domains do not undergo protease cleavage.
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.
CC -1- SIMILARITY: Contains 1 VWF_A domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AF021334; AAF21241.1; -.
DR HSSP; P11215; IBHO.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 3.
DR Pfam; PF00357; Integrin_alpha; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWF_A; 1.
DR Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;
KM Repeat; signal; Transmembrane.
FT SIGNAL 1 19 potential.
FT CHAIN 20 1161 Extracellular (Potential).
FT DOMAIN 1101 1121 Cytoplasmic (Potential).
FT REPEAT 1122 1161 FG-GAP 1.
FT REPEAT 86 87 FG-GAP 2.
FT DOMAIN 152 334 VWF_A.
FT REPEAT 352 402 FG-GAP 3.
FT REPEAT 403 454 FG-GAP 4.
FT REPEAT 456 517 FG-GAP 5.
FT REPEAT 519 577 FG-GAP 6.
FT REPEAT 582 634 FG-GAP 7.

```

FT CA\_BIND 467 475 Potential.  
 FT CA\_BIND 531 539 Potential.  
 FT CA\_BIND 594 602 Potential.  
 FT SITE 1126 1130 GFFR motif.  
 FT DISULFID 69 76 By similarity.  
 FT DISULFID 108 126 By similarity.  
 FT DISULFID 656 711 By similarity.  
 FT DISULFID 769 775 By similarity.  
 FT DISULFID 845 860 By similarity.  
 FT DISULFID 993 1017 By similarity.  
 FT DISULFID 1022 1027 By similarity.  
 FT CARBOHYD 61 61 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 89 89 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 245 245 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 393 393 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 696 696 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 734 734 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 784 784 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 907 907 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 936 936 N-linked (GlcNAc... ) (Potential).  
 FT CARBOHYD 1045 1045 N-linked (GlcNAc... ) (Potential).  
 SQ SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 15.4%; Score 138.5; DB 1; Length 1161;  
 Best Local Similarity 28.3%; Pred. No. 0.0032;  
 Matches 52; Conservative 36; Mismatches 71; Indels 25; Gaps 10;

QY 1 DLYVTLKSGSV-LHNNIEIYFVEQLAHKFIISPOLMSFVFSRGTTLMLKTE----- 54  
 DB 152 DIALFSGSGSINORDPAQKDFVQALMGEPASTSTLFLMYSNLIKTHFTTFEPKNIL 211  
 QY 55 DREQIRGELBELQKVLPGDVTYMEGFERASEQIYENRQGYRTA-SVTLALTDGELHED 113  
 DB 212 DPGSLVDPYVQLQ-----GLYVATGTGIRTWBELFHSNGSRKAKKILVITDGGQKXR 266  
 QY 114 LFFYSE--REANSRDIGAIVCVGV--FNE-TQLARI-----ADSKDHVPVNDGFOA 164  
 DB 267 PLRYSVDYIPADKA---GIIRYAIQVGDAPQEPALXELNTIGSAPQDHVFKVGN-FAA 322  
 QY 165 LQGI 168  
 DB 323 LRSL 326

## RESULT 8

Q04588 PRELIMINARY; PRT; 724 AA.

AC Q04588; (TREMURrel. 01, Created)  
 DT 01-NOV-1996 (TREMURrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TREMURrel. 26, Last annotation update)  
 DE Major antigen homologous sequence (empl00).  
 OS Eimeria maxima.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;  
 OC Eimeria.  
 NCBI\_TaxID=5804;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=3119203; PubMed=8426611; DOI=10.1016/0166-6851(93)90255-V;  
 RA Paramontes L.E., Hug D., Huembelin M., Weber G.;  
 RT "Sequence of a major Eimeria maxima antigen homologous to the Eimeria  
 tenella microneme protein Etp100.";  
 RT Mol. Biochem. Parasitol. 57:171-174(1993).  
 RL EMBL, M99058; AAA28076.1; -.  
 DR EMBL, M99058; AAA28076.1; -.  
 DR PIR, A48569; A48569.  
 DR HSSP, P07996; 1LSL.  
 DR GO, GO:0008083; F: growth factor activity; IEA.  
 DR InterPro, IPR000762; PTN MK.  
 DR InterPro, IPR000884; TSP1.  
 DR InterPro, IPR002035; VWF\_A.  
 DR Pfam, PF00090; TSP\_1; 5.  
 DR Pfam, PF00092; VMA; 1.  
 DR PRINTS, PR00453; VWFADOMAIN.

DR SMART, SM00209; TSP1; 6.  
 DR SMART, SM00327; VMA; 1.  
 DR PROSITE, PS50092; TSP1; 5.  
 DR PROSITE, PS50234; VMA; 1.  
 SQ SEQUENCE 724 AA; 75808 MW; AC2A0E7A346A7E3E CRC64;

Query Match 14.4%; Score 129.5; DB 2; Length 724;  
 Best Local Similarity 28.1%; Pred. No. 0.011;  
 Matches 48; Conservative 30; Mismatches 68; Indels 25; Gaps 10;

QY 1 DLYVTLKSGSV-LHNNIEIYFVEQLAHKFIISPOLMSFVFSRGTTLMLKTEORE 57  
 DB 47 DVLIVBESGSGISNTGKRSFTSNRAGTWPSPDDVRGVLVFGTSATVRMDLSRA 106  
 QY 58 QIRQELBELQKVLPGDVTYMEGFERASEQIYENRQGYRTA-SVTLALTDGELHED 113  
 DB 107 QMDLLAAAKKLPYAGSGTYTHGLAKA-BEILFSPQKGRDAPKILVMTDGA----- 161  
 QY 114 LFFYSEB-----ANRSRDIGAIVCVGV-KDENETQLARIA--DSKDHV 155  
 DB 162 -----SSRSQTLISAERKLRNGVILVILGVGTGVSASACRSIAGCDTSDTV 208

## RESULT 9

ITAM HUMAN STANDARD; PRT; 1152 AA.

AC P11215;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1) (Neutrophil adherence receptor).  
 GN Name: ITGAM; Synonyms: CD11B, CR3A;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88315033; PubMed=2457584;  
 RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;  
 RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B.";  
 RT J. Biol. Chem. 263:12403-12411(1988).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88190151; PubMed=2833753;  
 RA Arnout M.A., Remold-O'Donnell B., Pierce M.W., Harris P., Tenen D.G.;  
 RT "Molecular cloning of the alpha subunit of human and guinea pig leukocyte adhesion glycoprotein Mol. chromosomal localization and homology to the alpha subunits of integrins.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780(1988).  
 RL [3]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88257215; PubMed=2454931; DOI=10.1083/jcb.106.6.2153;  
 RA Arnout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;  
 RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mol (complement receptor type 3).";  
 RT J. Cell Biol. 106:2153-2158(1988).  
 RL [4]  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93123746; PubMed=8419480;  
 RA Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;  
 RT "Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-integrin gene family demonstrate remarkable conservation of genomic organization and suggest early diversification during evolution.";  
 RT J. Immunol. 150:480-490(1993).  
 RL [5]  
 RN [5]  
 RP SEQUENCE OF 9-1153 FROM N.A.  
 RX MEDLINE=89098893; PubMed=2563162;

RA Hickerein D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,  
RA Roth G.J.;  
RT "CDNA sequence for the alpha M subunit of the human neutrophil  
RT adherence receptor indicates homology to integrin alpha subunits.";   
RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).  
RN [6]  
RP SEQUENCE OF 1-9 FROM N.A.  
RX MEDLINE=92073318; PubMed=1683702;  
RA Shelley C.S., Arnaut M.A.;  
RT "The promoter of the CD1b gene directs myeloid-specific and  
RT developmentally regulated expression.";   
RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).  
RN [7]  
RP SEQUENCE OF 1-9 FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=92144986; PubMed=1346576;  
RA Paul H.L., Rosmarin A.G., Tegen D.G.;  
RT "Characterization of the myeloid-specific CD1b promoter.";   
RL Blood 79:865-870(1992).  
RN [8]  
RP SEQUENCE OF 17-31.  
RX MEDLINE=87076671; PubMed=3539202; DOI=10.1016/0167-4838(86)90037-3;  
RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnaut M.A.;  
RT "N-terminal sequence of human leukocyte glycoprotein Mol: conservation  
RT across species and homology to platelet IIb/IIIa.";   
RL Blochm. Biophys. Acta 874:368-371(1986).  
RN [9]  
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.  
RX MEDLINE=95174458; PubMed=7867070; DOI=10.1016/0092-8674(95)90517-0;  
RA Lee J.O., Rieu P., Arnaut M.A., Liddington R.C.;  
RT "Crystal structure of the A domain from the alpha subunit of integrin  
RT CR3 (CD11b/CD18).";   
RL Cell 80:631-638(1995).  
RN [10]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.  
RX MEDLINE=96363671; PubMed=8747460; DOI=10.1016/S0969-2126(01)00271-4;  
RA Lee J.O., Bankson L.A., Arnaut M.A., Liddington R.C.;  
RT "Two conformations of the integrin A-domain (1-domain): a pathway for  
RT activation?";   
RL Structure 3:1333-1340(1995).  
RN [11]  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.  
RX MEDLINE=98362599; PubMed=9687375; DOI=10.1016/S0969-2126(98)00093-8;  
RA Baldwin E.T., Server R.W., Bryant G.L., Jr., Curry K.A.,  
RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,  
RA Horton N.C., Kelley L.L., Milder A.M., Moon J.B., Mott J.E.,  
RA Matzner V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;  
RT "cation binding to the integrin CD11b I domain and activation model  
RT assessment.";   
RL Structure 6:923-935(1998).  
RN [12]  
RP 3D-STRUCTURE MODELING OF 17-616.  
RX MEDLINE=98226734; PubMed=9560195; DOI=10.1073/pnas.95.9.4870;  
RA Ouyig C., Springer T.A.;  
RT "Experimental support for a beta-propeller domain in integrin alpha-  
RT subunits and a calcium binding site on its lower surface.";   
RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).  
RN [13]  
RP FUNCTION: Integrin alpha-M/beta-2 is implicated in various  
RP adhesive interactions of monocytes, macrophages and granulocytes  
RP as well as in mediating the uptake of complement-coated particles.  
RP It is identical with CR-3, the receptor for the IC3b fragment of  
RP the third complement component. It probably recognizes the R-G-D  
RP peptide in C3b. Integrin alpha-M/beta-2 is also a receptor for  
RP fibrinogen, factor X and ICAM1. It recognizes P1 and P2 peptides  
RP of fibrinogen gamma chain.  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-M  
CC associates with beta-2.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- TISSUE SPECIFICITY: Predominantly expressed in monocytes and  
CC granulocytes.  
CC -1- DOMAIN: The integrin I-domain (insert) is a VWA domain. Integrins  
CC with I-domains do not undergo protease cleavage.  
CC -1- SIMILARITY: Belongs to the integrin alpha chain family.

CC -1- SIMILARITY: Contains 7 FG-GAP repeats.  
CC -1- SIMILARITY: Contains 1 VWA domain.  
CC DATABASE: NAME=PROW; NOTE=CD guide CD11b entry;  
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd11b.htm".  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL: J03925; AAAS9544.1; -;  
DR EMBL: M18044; AAAS9491.1; -;  
DR EMBL: J04145; AAAS9903.1; -;  
DR EMBL: S52227; AAB24821.1; -;  
DR EMBL: S52152; AAB24821.1; JOINED.  
DR EMBL: S52153; AAB24821.1; JOINED.  
DR EMBL: S52154; AAB24821.1; JOINED.  
DR EMBL: S52155; AAB24821.1; JOINED.  
DR EMBL: S52157; AAB24821.1; JOINED.  
DR EMBL: S52159; AAB24821.1; JOINED.  
DR EMBL: S52161; AAB24821.1; JOINED.  
DR EMBL: S52164; AAB24821.1; JOINED.  
DR EMBL: S52165; AAB24821.1; JOINED.  
DR EMBL: S52167; AAB24821.1; JOINED.  
DR EMBL: S52169; AAB24821.1; JOINED.  
DR EMBL: S52170; AAB24821.1; JOINED.  
DR EMBL: S52173; AAB24821.1; JOINED.  
DR EMBL: S52174; AAB24821.1; JOINED.  
DR EMBL: S52180; AAB24821.1; JOINED.  
DR EMBL: S52181; AAB24821.1; JOINED.  
DR EMBL: S52184; AAB24821.1; JOINED.  
DR EMBL: S52189; AAB24821.1; JOINED.  
DR EMBL: S52191; AAB24821.1; JOINED.  
DR EMBL: S52203; AAB24821.1; JOINED.  
DR EMBL: S52212; AAB24821.1; JOINED.  
DR EMBL: S52213; AAB24821.1; JOINED.  
DR EMBL: S52216; AAB24821.1; JOINED.  
DR EMBL: S52219; AAB24821.1; JOINED.  
DR EMBL: S52220; AAB24821.1; JOINED.  
DR EMBL: S52221; AAB24821.1; JOINED.  
DR EMBL: S52222; AAB24821.1; JOINED.  
DR EMBL: S52226; AAB24821.1; JOINED.  
DR EMBL: M76724; AAAS6410.1; -;  
DR EMBL: M64477; AAAS1960.1; -;  
DR PIR: A31108; RWHUIB.  
DR PDB: 1A8X; Model; @17-1152.  
DR PDB: 1BHO; X-ray; 1/2=-.  
DR PDB: 1BHO; X-ray; 1/2=-.  
DR PDB: 1IDN; X-ray; 1/2=-.  
DR PDB: 1IDO; X-ray; @140-331.  
DR PDB: 1JLM; X-ray; @143-334.  
DR PDB: 1MLU; X-ray; A=137-331.  
DR PDB: 1MF7; X-ray; A=144-337.  
DR PDB: 1MF7; X-ray; A=140-335.  
DR PDB: 1MA5; X-ray; A=144-345.  
DR Genew; HGNC:6149; ITGAM.  
DR MIM; 120980; -;  
DR GO; GO:0008305; C:integrin complex; TAS.  
DR GO; GO:0007155; P:cell adhesion; TAS.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF01859; FG-GAP; 3.  
DR Pfam; PF00357; Integrin\_alpha; 1.  
DR Pfam; PF00092; VWA; 1.  
DR PRINTS; PR01185; INTEGRINA.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00191; Int\_alpha; 5.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00242; INTEGRIN\_ALPHA; 1.



DR PROSITE; PSS0234; VMPA. 1.  
 KW 3D-structure; Calcium; Cell adhesion; Direct protein sequencing;  
 KW Glycoprotein; Integrin; Magnesium; Receptor; Repeat; Signal;  
 KW Transmembrane.  
 FT SIGNAL 1 16  
 FT CHAIN 17 1152 Integrin alpha-M.  
 Query Match 14.3%; Score 128.5; DB 1; Length 1152;  
 Best Local Similarity 26.7%; Pred. No. 0.024;  
 Matches 51; Conservative 30; Mismatches 63; Indels 39; Gaps 11;  
 QY 1 DLVPIIDKSGSVL-HHNNELVYFVEQLAHKIFISQQLMSFVSTRTTLMKTED---- 55  
 DB 150 DIAFLIDGSGSLLPHDRKMKPEVST-----VMEQLKSKTLFS-----LMQYSEPRRH 199  
 QY 56 -----REQIRQGLSELQKVLPGDPTMHGCFERASBOIYVENRGYRTA-SVITALT 106  
 DB 200 FTFRKPPNNNPISLVKPIQLL--GRTHRTATGRKRVRELFNTNARKAKAFILIVIT 257  
 QY 107 DSELHEDLPFYSE--REANRSRDGAIVCGVKNFNETOLAR-----IADS--KDHVP 157  
 DB 258 DGEKFGDPLGVEDVPEADRE---GVIRYVYGVDARFSEKSRQELMTLTKSKPRDHVFQ 314  
 QY 158 VNDGQALOGI 168  
 DB 315 VNN-PEALKTI 324  
 RESULT 10  
 CALC\_MOUSE STANDARD; PRT; 3119 AA.  
 ID 060847; P70322;  
 AC 15-JUL-1998 (Rel. 36, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Collagen alpha 1(XII) chain precursor.  
 GN Name=Col12a1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-1 AND XIIA-2).  
 RC STRAIN=C57BL/6J, and Swiss Webster; TISSUE=Skin;  
 RX MEDLINE=96170761; PubMed=8601036;  
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;  
 RT "Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development.";  
 RL Dev. Dyn. 204:432-445(1995).  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XIIA-2 AND XIIA-1).  
 RC STRAIN=C57BL/6J; TISSUE=Skin fibroblast;  
 RX MEDLINE=99348349; PubMed=10419532; DOI=10.1074/jbc.274.31.22053;  
 RA Kania A.M., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W., Olsen B.R., Nishimura I.;  
 RT "Structural variation of type XII collagen at its carboxyl-terminal NCI domain generated by tissue-specific alternative splicing.";  
 RL J. Biol. Chem. 274:22053-22059(1999).  
 CC -1- FUNCTION: Type XII collagen interacts with type I collagen-containing fibrils, the COL1 domain could be associated with the surface of the fibrils, and the COL2 and NC3 domains may be localized in the pericellular matrix (By similarity).  
 CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of nontriple-helical sequences (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoform=4;  
 CC Comment=The final tissue form of collagen XII may contain homotrimers or any combination of the various isoforms;  
 CC Name=XIIA-1;  
 CC IsoId=Q60847-1; Sequence=Displayed;

CC Name=XIIA-2; Synonym=ERK;  
 CC IsoId=Q60847-2; Sequence=VSP\_001151, VSP\_001152;  
 CC Name=XIIB-1;  
 CC IsoId=Q60847-3; Sequence=VSP\_001150;  
 CC Name=XIIB-2;  
 CC IsoId=Q60847-4; Sequence=VSP\_001150, VSP\_001151, VSP\_001152;  
 CC -1- TISSUE SPECIFICITY: Highest expression in tendons, perichondrium, skin, cornea, sclera, blood vessels, and peritoneum.  
 CC -1- DEVELOPMENTAL STAGE: The long NC3 XIIA isoforms are predominant at early stages (ED7 and 11); at later stages of development (ED15 and 17) the short NC3 XIIB forms become the major forms. As the short NC3 forms become the major product, the long splice variant continues to be expressed in several tissues, even after birth.  
 CC The long NC1 isoforms, XIIA-1 and XIIB-1, peak in 15-day old embryos and decrease in 17-day old ones. The expression of the short NC1 form XIIB-2 remains constant throughout late stages of embryonic development (ED15 and ED17).  
 CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at each end (By similarity).  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains (By similarity).  
 CC -1- PTM: O-glycosylation of isoform XIIA-2; glycosaminoglycan of chondroitin-sulfate type (By similarity).  
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with interrupted helices (FACIT) family.  
 CC -1- SIMILARITY: Contains 18 fibronectin type III domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -1- SIMILARITY: Contains 4 VMPA domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U25652; AAA99719.1; ALT\_SEQ.  
 DR EMBL; U57095; AAB07047.1; --  
 DR HSSP; P16614; LMRP.  
 DR WGD; MG1:88448; Col12a1.  
 DR Interpro; IPR008160; Collagen.  
 DR Interpro; IPR008985; Cona\_1like\_1ec\_g1.  
 DR Interpro; IPR003961; FN\_III.  
 DR Interpro; IPR008957; FN\_III-like.  
 DR Interpro; IPR003129; TSP\_N.  
 DR Interpro; IPR002035; VWF\_A.  
 DR Pfam; PF01391; Collagen; 5.  
 DR Pfam; PF00041; fn3; 18.  
 DR Pfam; PF02210; TSP\_N; 1.  
 DR Pfam; PR00092; VMA\_4.  
 DR PRINTS; PR00453; VWFADOMAIN.  
 DR SMART; SM00060; FN3; 18.  
 DR SMART; SM00210; TSPN; 1.  
 DR SMART; SM00327; VMA; 4.  
 DR PROSITE; PSS0853; FN3; 18.  
 DR PROSITE; PSS0234; VMPA; 4.  
 KW Alternative splicing; Cell adhesion; Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat; Signal; Structural protein.  
 FT SIGNAL 1 24  
 FT CHAIN 25 3119  
 FT DOMAIN 25 112 Collagen alpha 1(XII) chain.  
 FT DOMAIN 140 316 Fibronectin type-III 1.  
 FT DOMAIN 333 422 VMPA 1.  
 FT DOMAIN 333 422 Fibronectin type-III 2.  
 FT DOMAIN 444 620 VMPA 2.  
 FT DOMAIN 635 723 Fibronectin type-III 3.  
 FT DOMAIN 726 813 Fibronectin type-III 4.  
 FT DOMAIN 817 905 Fibronectin type-III 5.  
 FT DOMAIN 908 997 Fibronectin type-III 6.  
 FT DOMAIN 999 1087 Fibronectin type-III 7.  
 FT DOMAIN 1090 1179 Fibronectin type-III 8.  
 FT DOMAIN 1203 1375 VMPA 3.

FT	DOMAIN	1388	1476	Fibronectin type-III 9.
FT	DOMAIN	1478	1567	Fibronectin type-III 10.
FT	DOMAIN	1569	1656	Fibronectin type-III 11.
FT	DOMAIN	1660	1747	Fibronectin type-III 12.
FT	DOMAIN	1758	1847	Fibronectin type-III 13.
FT	DOMAIN	1849	1937	Fibronectin type-III 14.
FT	DOMAIN	1939	2028	Fibronectin type-III 15.
FT	DOMAIN	2030	2119	Fibronectin type-III 16.
FT	DOMAIN	2121	2208	Fibronectin type-III 17.
FT	DOMAIN	2212	2298	Fibronectin type-III 18.
FT	DOMAIN	2329	2501	WFA 4.
FT	DOMAIN	2525	2717	GSP N-terminal.
FT	DOMAIN	2456	2751	Triple-helical region (COL2) with 1 imperfection.
FT	DOMAIN	2752	2839	Nonhelical region (NC3).
FT	DOMAIN	2900	2942	Nonhelical region (NC2).
FT	DOMAIN	2943	3045	Triple-helical region (COL1) with 2 imperfections.
FT	DOMAIN	3046	3119	Nonhelical region (NC1).
FT	SITE	866	868	Cell attachment site (Potential).
FT	SITE	2784	2786	Cell attachment site (Potential).
FT	SITE	2896	2898	Cell attachment site (Potential).
FT	MOD_RES	2945	2945	Hydroxyproline (By similarity).
FT	MOD_RES	2948	2948	Hydroxyproline (By similarity).
FT	MOD_RES	2951	2951	Hydroxyproline (By similarity).
FT	MOD_RES	2960	2960	Hydroxyproline (By similarity).
FT	MOD_RES	2966	2966	Hydroxyproline (By similarity).
FT	MOD_RES	2969	2969	Hydroxyproline (By similarity).
FT	MOD_RES	2972	2972	Hydroxyproline (By similarity).
FT	MOD_RES	2984	2984	Hydroxyproline (By similarity).
FT	MOD_RES	3001	3001	Hydroxyproline (By similarity).
FT	MOD_RES	3004	3004	Hydroxyproline (By similarity).
FT	MOD_RES	3015	3015	Hydroxyproline (By similarity).
FT	MOD_RES	3024	3024	Hydroxyproline (By similarity).
FT	MOD_RES	3027	3027	Hydroxyproline (By similarity).
FT	MOD_RES	3030	3030	Hydroxyproline (By similarity).
FT	DOMAIN	869	872	Poly-Thr.
FT	CARBOHYD	704	704	N-linked (GlcNAc . . ) (Potential).
FT	CARBOHYD	802	802	O-linked (XyL . . ) (chondroitin sulfate)
FT	CARBOHYD	893	893	O-linked (XyL . . ) (chondroitin sulfate)
FT	CARBOHYD	985	985	O-linked (XyL . . ) (chondroitin sulfate)
FT	CARBOHYD	1769	1769	N-linked (GlcNAc . . ) (Potential).
FT	CARBOHYD	2212	2212	N-linked (GlcNAc . . ) (Potential).
FT	CARBOHYD	2533	2533	N-linked (GlcNAc . . ) (Potential).
FT	CARBOHYD	2684	2684	N-linked (GlcNAc . . ) (Potential).
FT	VARSPLIC	25	1190	Missing (in isoform XIIB-1 and isoform XIIB-2).
FT	VARSPLIC	3062	3064	/FTId=VSP 001150.
FT	VARSPLIC	3065	3119	BEY -> GSG (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001151.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001152.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001153.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001154.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001155.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001156.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001157.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001158.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001159.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001160.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001161.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001162.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001163.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001164.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001165.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001166.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001167.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001168.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001169.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001170.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001171.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001172.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001173.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001174.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001175.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001176.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001177.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001178.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001179.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001180.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001181.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001182.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001183.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001184.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001185.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001186.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001187.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001188.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001189.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001190.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001191.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001192.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001193.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001194.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001195.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001196.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001197.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001198.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001199.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001200.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001201.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001202.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001203.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001204.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001205.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001206.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001207.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001208.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001209.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001210.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001211.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001212.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001213.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001214.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001215.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001216.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001217.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001218.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001219.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001220.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001221.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001222.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001223.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001224.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001225.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001226.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001227.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001228.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001229.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001230.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001231.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001232.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001233.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001234.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001235.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001236.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001237.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001238.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001239.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001240.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001241.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001242.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001243.
FT	VARSPLIC	3065	3119	Missing (in isoform XIIA-2 and isoform XIIB-2).
FT	VARSPLIC	3065	3119	/FTId=VSP 001244.
FT	VARSPLIC	3065	3119	

Db	557	SSDAE----	RDPALIKRNSDVELFANGVADNARSELEAIAISPPAEHTHVTFED-PDAFORI	612
 RESULT 11				
ID	Q8T6U5	PRELIMINARY;	PRT; 441 AA.	
AC	Q8T6U5			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DE	Proximal thred matrix protein 1 variant a.			
OS	Myltilus edulis (Blue mussel).			
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloidea;			
CC	Mytiloidea; Mytilidae; Mytilus.			
NCBI_TaxID=6550;	[1]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;			
RX	Sun C., Lucas J.M., Waite J.H.;			
RA	"Collagen-binding matrix proteins from elastomeric extraorganismic			
RT	bysael fibers.";			
RL	Biomacromolecules 3:1240-1248(2002).			
DR	EMBL; AF414454; AAL83537.1; -			
DR	GO; GO:0005198; F.structural molecule activity; IEA.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF000092; VMA; 2.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00327; VMA; 2.			
DR	PROSITE; PS50234; VWFA; 2.			
KW	Matrix protein.			
SQ	SEQUENCE 441 AA; 47543 MW; 881DBBD36B891D2B CRC64;			
 Query Match 13.9%; Score 125; DB 2; Length 441;				
Best Local Similarity 24.7%; Pred. NO. 0.016;				
Matches	47;	Conservative	44;	Mismatches 73; Indels 26; Gaps 12
QY	1	DLYFLDLSGSV----	LHNNREIYYFVQLAHKF--ISPO-LRMSPFYSTRTGTLMLKT	53
DB	241	DIAPFDSDASSINANNPNNVKMDFKMDIVDFRNKTPGDGTGFVVTFADATKQFGLK	300	
QY	54	E-DREOIRQGIEBELQKLPG--GPTVMHEGPERSAQIYYENRG--YRPAHYIALT	106	
DB	301	DYSSKAELTKGAID---KVTPSIIGQTAIQGLGNARKLEV-FPNNRGGRREVQKVILLT	356	
QY	107	DGETL-HEDLFYSEREANRSRDICAIYYCVGV-KDFMETQOLARIADSKDHVPVNDGFO	163	
DB	357	DGQNNNGHS-----PEHESSLRKSGSVIVALVGVGIFLKSLELINIASSEEVNF-TTSSEN	411	
QY	164	ALQGIHSIL 173		
DB	412	KLSKIMENVV 421		
 RESULT 12				
ID	Q8T5C3	PRELIMINARY;	PRT; 444 AA.	
AC	Q8T5C3;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2003 (TREMBLrel. 24, Last annotation update)			
OS	Proximal thread matrix protein 1b.			
OS	Mytilus edulis (Blue mussel).			
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Mytiloidea;			
CC	Mytiloidea; Mytilidae; Mytilus.			
NCBI_TaxID=6550;	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=22313036; PubMed=12425661; DOI=10.1021/bm0255903;			
RA	Sun C., Lucas J.M., Waite J.H.;			
RT	"Collagen-binding matrix proteins from elastomeric extraorganismic			
RT	bysael fibers.";			
RL	Biomacromolecules 3:1240-1248(2002).			
DR	EMBL; AY053390; AAL17973.1; -			

DR GO:0005198; Functional molecule activity; IEA.  
 DR InterPro: IPR002035; VMA\_A.  
 DR Pfam: PF00092; VMA; 2.  
 DR PRINTS: PR00453; VMAFADOMAIN.  
 DR SMART: SM00327; VMA; 2.  
 DR PROSITE: PSS0234; VMA; 2.  
 KM Matrix protein.  
 SQ SEQUENCE 444 AA; 4781 MW; D2C605347450C931 CRC64;  
 Query Match 13.9%; Score 125; DB 2; Length 444;  
 Best local similarity 24.7%; Pred. No. 0.016;  
 Matches 47; Conservative 44; Mismatches 73; Indels 26; Gaps 12;  
 QY 1 DLYFIDKSGSV---LHMNEIYFVEQLAHKF--ISPO-LRMSFVFSFGTTLAKLT 53  
 DB 244 DIAFVFPASSSINNANNNNYGLMKDFMKDIYDRNKTGPDGTGPAVATPDRATKQGLK 303  
 QY 54 E--DREIDRGLELQVLTG--GDYMHGSPFASQIYENQG--YRTASVITALT 106  
 DB 304 DYSSKAEIKGAID---KVTFSIIIGQTAIGDLLENARLEV-FPNNMGGRERVKVILLT 359  
 QY 107 DGEI--HEDLFPYSESRANRSDGATVCGV-KDPEHQIARIADSKHVPVNDGQ 163  
 DB 360 DQGNHKS---PEHSSLRKEGVVAIGVGTGLKSLINIASSEVTF-TTSEFN 414  
 QY 164 ALQGIHSIL 173  
 DB 415 KLSKIMENVV 424  
 RESULT 13  
 CALC CHICK STANDARD; PRT; 3124 AA.  
 ID CA13944; 004509;  
 AC P13944; 004509 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE COLLAGEN alpha 1(XII) chain precursor (Fibrochimerin).  
 GN Name=COL12A1;  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=White Leghorn;  
 RX MEDLINE=92011862; PubMed=1918137; DOI=10.1083/jcb.115.1.209;  
 RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,  
 RA Nishida Y., Ohara M., Kimata K.;  
 RT "The complete primary structure of type XII collagen shows a chimeric  
 RT molecule with reiterated fibronectin type III motifs, von Willebrand  
 RT factor A motifs, a domain homologous to a noncollagenous region of  
 RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp  
 RT site."  
 RT J. Cell Biol. 115:209-221(1991).  
 RN [2]  
 RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND  
 RP 2846-2873.  
 RX MEDLINE=90062079; PubMed=2584192;  
 RA Gordon M.K., Gerecke D.R., Dubler B., van der Reest M., Olsen B.R.;  
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;  
 RT "Type XII collagen: distinct extracellular matrix component discovered  
 RT by cDNA cloning."  
 RT Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044(1987).  
 RN [4]  
 RP SEQUENCE OF 1-1283 FROM N.A. (ISOFORM SHORT), AND ALTERNATIVE

RP SPLICING.  
 RC TISSUE=Embryo;  
 RX MEDLINE=93042014; PubMed=1420368; DOI=10.1016/0167-4781(92)90145-B;  
 RA Trueb J., Trueb B.;  
 RT "The two splice variants of collagen XII share a common 5' end."  
 RL Biochim. Biophys. Acta 1171:97-98(1992).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=95370352; PubMed=7642694; DOI=10.1083/jcb.130.4.1005;  
 RA Koch M., Bohmann B., Matchison M., Hagios C., Trueb B., Chiquet M.;  
 RT "Large and small splice variants of collagen XII: differential  
 RT expression and ligand binding."  
 RL J. Cell Biol. 130:1005-1014(1995).  
 CC -1- FUNCTION: Type XII collagen interacts with type I collagen-  
 CC containing fibrils, the COL1 domain could be associated with the  
 CC surface of the fibrils, and the COL2 and NC3 domains may be  
 CC localized in the pericellular matrix.  
 CC -1- SUBUNIT: Trimer of identical chains each containing 190 kDa of  
 CC nontriple-helical sequences.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=The final tissue form of collagen XII may contain  
 CC homotrimers of either isoform Long or isoform Short or any  
 CC combination of isoform Long and isoform Short. Only isoform Long  
 CC is a proteoglycan. Isoform Long has more restricted expression  
 CC in embryonic tissue than isoform Short;  
 CC Name=Long;  
 CC IsoId=P13944-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P13944-2; Sequence=VSP\_001148;  
 CC -1- TISSUE SPECIFICITY: Type XII collagen is present in tendons,  
 CC ligaments, perichondrium, and perosteum, all dense connective  
 CC tissues containing type I collagen.  
 CC -1- DOMAIN: This sequence defines five distinct domains, two triple-  
 CC helical domains (COL1 and COL2) and three nontriple-helical  
 CC domains (NC1, NC2, and NC3).  
 CC -1- PTM: The triple-helical tail is stabilized by disulfide bonds at  
 CC each end.  
 CC -1- PTM: Prolines at the third position of the tripeptide repeating  
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
 CC -1- PTM: O-glycosylated; glycosaminoglycan of chondroitin-sulfate type  
 CC (by similarity).  
 CC -1- SIMILARITY: Belongs to the fibril-associated collagens with  
 CC interrupted helices (FACIT) family.  
 CC -1- SIMILARITY: Contains 18 fibronectin type III domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -1- SIMILARITY: Contains 4 VMA domains.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D00824; BAA00701.1; -  
 DR EMBL: X61024; CAA43358.1; -  
 DR EMBL: M17375; AAA48718.1; -  
 DR EMBL: J05137; AAA48635.1; -  
 DR EMBL: X67327; CAA47744.1; -  
 DR PIR: A40020; A40020.  
 DR HSP: P56199; 10C5.  
 DR InterPro: IPR008160; Collagen.  
 DR InterPro: IPR008985; Cona\_like\_1ec\_g1.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR008957; FN\_III-like.  
 DR InterPro: IPR003129; TSP\_N.  
 DR InterPro: IPR002035; VMA\_A.  
 DR Pfam: PF01391; Collagen; 4.  
 DR Pfam: PF00041; Fn3; 17.  
 DR Pfam: PF02210; TSP\_N; 1.  
 DR Pfam: PF00092; VMA; 4.



DT 16-OCT-2001 (Ref. 40, last sequence update)  
 DT 05-JUL-2004 (Ref. 44, last annotation update)  
 DE Integrin alpha-D precursor (leukointegrin alpha D) (CD11d) (ADB2).  
 GN Name:ITGAD;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Spleen;  
 RX MEDLINE=66111955; PubMed=8777714; DOI=10.1016/1074-7613(95)90058-6;  
 RA Van der Vlieten M., Le Trong H., Wood C.V., Moore P.F., St John T.,  
 RA Staunton D.E., Gallatin W.M.;  
 RT "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-  
 RT 3.";  
 RL Immunity 3:683-690(1995).  
 RN [2]  
 RP SEQUENCE OF 1-235 FROM N.A.  
 RX MEDLINE=20187620; PubMed=10722744; DOI=10.1074/jbc.275.12.8959;  
 RA Noci J.D., Johnson A.K., Dillon J.D.;  
 RT "Structural and functional characterization of the leukocyte integrin  
 RT gene CD1d. Essential role of Sps and Sps3.";  
 RL J Biol. Chem. 275:8959-8969(2000).  
 RN [3]  
 RP SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.  
 RX MEDLINE=96257236; PubMed=8666289; DOI=10.1016/0378-1119(95)00869-1;  
 RA Wong D.A., Davis E.M., LeBeau M., Springer T.A.;  
 RT "Cloning and chromosomal localization of a novel gene-encoding a human  
 RT beta 2-integrin alpha subunit.";  
 RL Gene 171:291-294(1996).  
 RN [4]  
 RP INTERACTION WITH VCAM1.  
 RX MEDLINE=99059842; PubMed=9841932;  
 RA Grayson M.H., Van der Vlieten M., Sterbinsky S.A., Michael Gallatin W.,  
 RA Hoffman P.A., Staunton D.E., Bochner B.S.;  
 RT "Alphadelta2 integrin is expressed on human eosinophils and functions  
 RT as an alternative ligand for vascular cell adhesion molecule 1 (VCAM-  
 RT 1).";  
 RL J. Exp. Med. 188:2187-2191(1998).  
 RN [5]  
 RP INTERACTION WITH VCAM1.  
 RX MEDLINE=99370002; PubMed=10438935;  
 RA Van der Vlieten M., Crowe D.T., Hoekstra D., Vazquez R., Hoffman P.A.,  
 RA Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;  
 RT "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a  
 RT binding interface between I domain and VCAM-1.";  
 RL J. Immunol. 163:1984-1990(1999).  
 RN [6]  
 RP FUNCTION: Integrin alpha-D/beta-2 is a receptor for ICAM3 and  
 CC VCAM1. May play a role in the atherosclerotic process such as  
 CC clearing lipoproteins from plaques and in phagocytosis of blood-  
 CC borne pathogens, particulate matter, and senescent erythrocytes  
 CC from the blood.  
 CC [7]  
 RP SUBUNIT: Heterodimer of an alpha and a beta subunit. Alpha-D  
 CC associates with beta-2.  
 CC [8]  
 RP SUBCELLULAR LOCATION: Type I membrane protein.  
 CC [9]  
 RP TISSUE SPECIFICITY: Expressed moderately on myelomonocytic cell  
 CC lines and subsets of peripheral blood leukocytes and strongly on  
 CC tissue-specialized cells, including macrophages foam cells within  
 CC atherosclerotic plaques, and on splenic red pulp macrophages.  
 CC [10]  
 RP DOMAIN: The integrin I-domain (insert) is a VWA domain. Integrins  
 CC with I-domains do not undergo protease cleavage.  
 CC [11]  
 RP SIMILARITY: Belongs to the integrin alpha chain family.  
 CC [12]  
 RP SIMILARITY: Contains 7 FG-GAP repeats.  
 CC [13]  
 RP SIMILARITY: Contains 1 VWA domain.  
 CC [14]  
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: U37028; AAB38547.1; -;  
 DR EMBL: U40274; AAB60634.1; -;  
 DR EMBL: U40275; AAB60635.1; -;  
 DR EMBL: U40276; AAB60636.1; -;  
 DR EMBL: U40277; AAB60637.1; -;  
 DR EMBL: U40279; AAB60638.1; -;  
 DR EMBL: U40278; AAB60638.1; JOINED.  
 DR EMBL: AF187881; AAF62875.1; -;  
 DR HSSP: P11215; IBHQ.  
 DR GENE: HGNC:6146; ITGAD.  
 DR MIM: 602453; -;  
 DR GO: GO:0008305; C:integrin complex; TAS.  
 DR GO: GO:0016337; P:cell-cell adhesion; NAS.  
 DR GO: GO:0007160; P:cell-matrix adhesion; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR000413; Integrin\_alpha.  
 DR InterPro: IPR02035; VWF\_A.  
 DR Pfam: PF01839; FG-GAP; 3.  
 DR Pfam: PF00357; Integrin\_alpha; 1.  
 DR Pfam: PF00092; VWA; 1.  
 DR PRINTS: PRO1185; INTEGRINA.  
 DR PRINTS: PRO0453; VWFADOMAIN.  
 DR SMART: SM00191; Int\_alpha; 5.  
 DR SMART: SM00327; VWA; 1.  
 DR PROSITE: PS00242; INTEGRIN\_ALPHA; 1.  
 DR PROSITE: PS50234; VWA; 1.  
 KW Calcium; Cell adhesion; Glycoprotein; Integrin; Magnesium; Receptor;  
 KW Repeat; Signal; Transmembrane  
 FT SIGNAL 1 17  
 FT CHAIN 18 1162  
 FT DOMAIN 18 1100  
 FT TRANSMEM 1101 1121  
 FT DOMAIN 1122 1162  
 FT REPEAT 32 85  
 FT REPEAT 86 7  
 FT DOMAIN 150 332  
 FT REPEAT 350 400  
 FT REPEAT 401 452  
 FT REPEAT 454 516  
 FT REPEAT 518 576  
 FT REPEAT 581 633  
 FT CA\_BIND 465 478  
 FT CA\_BIND 530 538  
 FT CA\_BIND 593 601  
 FT SITE 1127 1131  
 FT DISULFID 67 74  
 FT DISULFID 106 124  
 FT DISULFID 655 710  
 FT DISULFID 769 775  
 FT DISULFID 846 861  
 FT DISULFID 994 1018  
 FT DISULFID 1023 1028  
 FT CARBOHYD 59 59  
 FT CARBOHYD 87 87  
 FT CARBOHYD 99 99  
 FT CARBOHYD 391 391  
 FT CARBOHYD 691 691  
 FT CARBOHYD 733 733  
 FT CARBOHYD 873 873  
 FT CARBOHYD 957 957  
 FT CARBOHYD 1046 1046  
 FT CONFLICT 500 500  
 FT CONFLICT 515 518  
 FT CONFLICT 825 825  
 FT CONFLICT 984 984  
 FT SEQUENCE 1162 AA; 126885 MW; F236A1A3545D77D CRC64;  
 Query Match 13.6%; Score 122.5; DB 1; Length 1162;  
 Best Local Similarity 26.0%; Pred. No. 0.08; Indels 19; Gaps 7;  
 Matches 47; Conservative 35; Mismatches 80;  
 1 DLYEILDKSGSV-LHHNIEIYFVEQLAHKFIPLQRLMSFIIVFSTRGTTLMKLTEDRQI 59

```

Db      150  DIYFLIDSGSIDQNDPNQKGFVQAVMGQPEGTDTLFALMQY8---NLKIHFTPTQF 205
QY      60  RQGLEELQKVL P---GADTYMHEGFERASEQIYYENRQGYRTA-SVIALTDGELHEDL 114
Db      206  RTSPTSQOSLVDPIVOLKGLFTATGILTVVTLPHHNGAKSAKILITVTDGQKYXDP 265
QY      115  FYSEREARNSRDIGALVYCVGVKDFNETQLAR-----IADSKDHVPFVNDGFQALQG 167
Db      266  LEYSD-VIPOAEKAGILRYAIGVGHAFQGPTRQELNTISSAPPQDHVFKV-DNFALGS 323
QY      168  I 168
Db      324  I 324

```

Search completed: June 13, 2005, 20:03:16  
 Job time : 59.0753 secs